PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 24 JUL 2003
Patent Office PCT
Canberra

I, JULIE BILLINGSLEY, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PS 3339 for a patent by THE AUSTRALIAN NATIONAL UNIVERSITY as filed on 02 July 2002.

WITNESS my hand this
Fourteenth day of July 2003

J. Billingsley

JULIE BILLINGSLEY
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

The Australian National University

PROVISIONAL SPECIFICATION

Invention Title:

Method of producing plants having enhanced transpiration efficiency and plants produced therefrom I

The invention is described in the following statement:

FIELD OF THE INVENTION

The present invention relates to the field of plant breeding and the production of genetically engineered plants. More specifically, the invention described herein provides genes that are capable of enhancing the transpiration efficiency of a plant when expressed therein. These genes are particularly useful for the production of plants having enhanced transpiration efficiency, by both traditional plant breeding and genetic engineering approaches. The invention further extends to plants produced by the methods described herein.

BACKGROUND TO THE INVENTION

1. General

This specification contains nucleotide and amino acid sequence information prepared using PatentIn Version 3.1, presented herein after the claims. Each nucleotide sequence is identified in the sequence listing by the numeric indicator <210> followed by the sequence identifier (e.g. <210>1, <210>2, etc). The length and type of sequence (DNA, protein (PRT), etc), and source organism for each nucleotide sequence, are indicated by information provided in the numeric indicator fields <211>, <212> and <213>, respectively. Nucleotide sequences referred to in the specification are defined by the term "SEQ ID NO:", followed by the sequence identifier (eg. SEQ ID NO: 1 refers to the sequence in the sequence listing designated as <400>1).

The designation of nucleotide residues referred to herein are those recommended by the IUPAC-IUB Biochemical Nomenclature Commission, wherein A represents Adenine, C represents Cytosine, G represents Guanine, T represents thymine, Y represents a pyrimidine residue, R represents a purine residue, M represents Adenine or Cytosine, K represents Guanine or Thymine, S represents Guanine or Cytosine, W represents Adenine or Thymine, H represents a nucleotide other than Guanine, B represents a nucleotide other than Adenine, V represents a nucleotide other than Thymine, D represents a nucleotide other than Cytosine and N represents any nucleotide residue.

As used herein the term "derived from" shall be taken to indicate that a specified integer is obtained from a particular source albeit not necessarily directly from that source.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated step or element or integer or group of steps or elements or integers but not the exclusion of any other step or element or integer or group of elements or integers.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations or any two or more of said steps or features.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended for the purposes of exemplification only. Functionally equivalent products, compositions and methods are clearly within the scope of the invention, as described herein.

2. Description of the related art

It is well known that virtually all plants require a certain quantity of water for proper growth and development, because CO_2 fixation and photosynthate assimilation by plants cost water. A significant quantity of water absorbed by plants from the soil returns to the atmosphere *via* plant transpiration.

Transpiration efficiency is a measure of the amount of dry matter produced by a plant per unit of water transpired, or, in other words, carbon gain relative to water lost through transpiration.

For plants having low transpiration efficiency, or when water is in short supply, the loss of water through transpiration can limit key metabolic processes associated with

plant growth and development. For example, during drought, or when plants having low transpiration efficiency are grown in arid and semi-arid environments, plant productivity as determined by dry matter production or photosynthetic rate, is considerably reduced. Accordingly, the production of plants having enhanced water use efficiency or transpiration efficiency is highly desirable for their adaptation to arid or semi-arid conditions, or to enhance their drought resistance.

The enhancement of water use efficiency or transpiration efficiency by plants is also highly desirable in consideration of global climatic change and increasing pressure on world water resources. The inefficient utilization of agricultural water is known to impact adversely upon the supply of navigable water, potable water, and water for industrial or recreational use. Accordingly, the production of plants having enhanced transpiration efficiency is highly desirable for reducing the pressure on these water resources. It is also desirable for increasing plant productivity under well-watered conditions.

By enhancing transpiration efficiency, carbon gain rates are enhanced per unit of water transpired, thereby stimulating plant growth under well-watered conditions, or alternatively, under mild or severe drought conditions. This is achieved by enhancing carbon gain more than transpiration rate, or by reducing the amount of water lost at any particular rate of carbon fixation. Those skilled in the art also consider that for a given growth rate plants having enhanced transpiration efficiency dry out soils more slowly, and use less water, than less efficient near-isogenic plants.

Several chemical as well as environmental pre-treatments have been described for enhancing the ability of plant seedlings to survive drought, either by reducing transpiration or by reducing the amount of water that is actually lost to the atmosphere.

Known environmental treatments largely involve the use of physical barriers. Whilst placing a physical barrier over plant stomata is known to reduce water loss via transpiration, the procedure is not always desirable or practicable for field-grown crops. For example, physical barriers over plant stomata may inhibit certain gas-exchange processes of the plant. It is more desirable to enhance actual transpiration

efficiency or water use efficiency of the plant through manipulation of intrinsic plant function.

Chemical agents are typically the so-called "anti-transpirant" or "anti-desiccant" agents, both of which are applied to the leaves. Anti-transpirants are typically films or metabolic anti-transpirants.

These products form a film on leaves, thereby either blocking stomatal pores, or coating leaf epidermal cells with a water-proof film. Typical film anti-transpirants include waxes, wax-oil emulsions, higher alcohols, silicones, plastics, latexes and resins. For example, Elmore, United States Patent No. 4,645,682 disclosed an anti-transpirant consisting of an aqueous paste wax; Cushman *et al.*, United States Patent Nos. 3,791,839 and 3,847,641 also disclosed wax emulsions for controlling transpiration in plants; and Petrucco *et al.*, United States Patent No. 3,826,671, disclosed a polymer composition said to be effective for controlling transpiration in plants.

Metabolic anti-transpirants generally close stomata, thereby reducing the rate of transpiration. Typical metabolic anti-transpirants include succinic acids, phenylmercuric acetate, hydroxysulfonates, the herbicide atrazine, sodium azide, and phenylhydrazones, as well as carbon cyanide.

Compounds having plant growth regulator activity have also been shown to be useful for reducing transpiration. For example, Bliesner *et al.*, United States Patent No. 4,671,816, disclosed an acetylene compound, said to possess utility for regulating plant growth, whilst Kuznetsov *et al.* (Russian Patent No. SU 1,282,492; ., Russian Patent Application No. SU 1,253,559-A1), and Smirnov *et al* (Russian Patent No. SU 1,098,934) disclosed the use of derivatives of 2-methyl-5-hydroxybenzimidazole, and the chloride or bromide salts thereof, as anti-transpirant growth regulators. Vichnevetskaia (USSN 5,589,437 issued December 31, 1996) also describe hydroxybenzimidazole derivatives for enhancing the drought resistance of plants by reducing transpiration, however have the advantage of being applicable to plant seed or roots. Schulz *et al.*, United States Patent No. 4,943,315, also disclosed formulations comprising an acetylene and a phenylbenzylurea compound, for

reducing transpiration in plants and/or for avoiding impairment to plants caused by heat and dry conditions. Abscisic acid has also been shown to reduce or suppress transpiration in plants (eg. *Helv. Chim. Acta, 71*, 931, 1988; *J. Org. Chem., 54*, 681, 1989; and Japanese Patent Publication No. 184,966/1991).

Metabolic anti-transpirants are costly to produce and often exhibit phytotoxic effects or inhibit plant growth Kozlowski (1979), *In:* Tree Growth and Environmental Stresses (Univ. of Washington Press, Seattle and London), and are not practically used.

Recent studies have examined alternative methods for enhancing transpiration efficiency, particularly breeding approaches to select lines that grow more efficiently under mild drought conditions. Carbon isotope discrimination has been used to identify Arabidopsis ecotypes with contrasted transpiration efficiencies (Masle *et al., In:* Stable isotopes and plant carbon-water relations, Acad. Press, Physiol. Ser., pp371-386, 1993) and to assist conventional breeding of new plant varieties in a number of species (Hall *et al., Plant Breeding Reviews 4*, 81-113, 1994) including rice (Farquhar *et al., In:* Breaking the Yield Barrier, ed KG Cassman, IRRI, 95, 101) and most recently wheat (Rebetzke et al. Crop Science 42:739-745, 2002).

No single gene has been identified as being capable of enhancing transpiration efficiency when expressed *in planta*. Transpiration efficiency may well be multigenic. As a consequence, the genes and signalling pathways that regulate the photosynthetic and/or stomatal components of the transpiration efficiency mechanism in plants have not been identified or characterized.

Moreover, notwithstanding that the effect of down-regulating expression of the *Rubisco* gene, or mutation in genes involved in abscisic acid (eg. *aba, abi*), are known to modify transpiration efficiency to some extent through stomatal closure, the consequence of such modifications is not totally specific, resulting in pleiotropic effects.

Arabidopsis thaliana ecotype Landsberg *erecta* (L-*er*) is one of the most popular ecotypes and is used widely for both molecular and genetic studies. It harbors the *erecta* (*er*) mutation, which confers a compact inflorescence, blunt fruits, and short

petioles. There are a number of *er* mutant alleles. Phenotypic characterization of the mutant alleles suggests a role for the wild type *ER* gene in regulating plant morphogenesis, particularly the shapes of organs that originate from the shoot apical meristem. Torii *et al., The Plant Cell 8*, 735, 1996, showed that the *ER* gene encodes a putative receptor protein kinase comprising a cytoplasmic protein kinase catalytic domain, a transmembrane region, and an extracellular domain consisting of leucine-rich repeats, which are thought to interact with other macromolecules.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a is a graphical representation showing the CO_2 assimilation rates (μmol C m^2 s^{-1}) of several genotypes of *A. thaliana*. The genotypes of plants are indicated on the x-axis, and CO_2 assimilation rates indicated on the ordinate. Col indicates a genetic background of the ecotype Columbia. Ld indicates a genetic background of the ecotype Landsberg. Plants expressing wild type *ERECTA* alleles were either in a Col (Col4-*ER*) or Ld (Ld-*ER*) background. Plants that were homozygous for a mutant *er* allele were either in a Col background (Col-*er105* or Col-*er2*) or a Ld background (Ld-*er1*). Plants designated as F1 (Col-*ER* x Ld-*er*) were heterozygous *ER/er1*. Data indicate that, in a Col background, the *er105* mutation leads to reduced CO_2 assimilation rate, whilst the *er1* mutation enhances CO_2 assimilation rate in a Ld background.

Figure 1b is a graphical representation showing the stomatal conductance (mol H_20 m^2 s^{-1}) of several genotypes of *A. thaliana*. The genotypes of plants are indicated on the x-axis and are the same as described in the legend to Figure 1a. Stomatal conductances are indicated on the ordinate. Data indicate that, in a Col background, the *er2* mutation significantly enhances stomatal conductance, whilst the *er1* mutation significantly enhances stomatal conductance in a Ld background.

Figure 1c is a graphical representation showing the transpiration efficiency of (mmol C mol H_20^{-1}) of several genotypes of *A. thaliana*, as determined by the ratio of CO_2 assimilation rate to stomatal conductance. The genotypes of plants are indicated on the x-axis and are the same as described in the legend to Figure 1a. Transpiration efficiency is indicated on the ordinate. Data indicate that transpiration efficiency is enhanced in plants expressing a wild type *ER* allele relative to a mutant *er* allele, in

both Ld and Col backgrounds. The lowest transpiration efficiency was observed for plants that are homozygous for the *er105* allele (ie. Col-*er105*), consistent with the fact that this allele disrupts *ERECTA* expression. From the data in Figures 1a-1c, it is apparent that the lower transpiration efficiency of plants expressing the *er105* allele is largely due to a reduced CO_2 fixation rate, whereas for both the *er2* and *er1* alleles, reduced transpiration efficiency is largely due to enhanced stomatal conductance. The transpiration efficiency of the F1 heterozygote plant was intermediate between the transpiration efficiencies of its parents, suggesting codominance of these alleles. The F1, however, had a transpiration efficiency closer to that of the pollen donor parent, Ld-*er1*.

Figure 2a is a graphical representation showing the stomatal densities (Number of stomata mm^{-2} leaf) for several genotypes of *A. thaliana* in three independent experiments. The genetic backgrounds of plants are indicated on the x-axis (Col, Columbia; Ld, Landsberg), and stomatal densities are indicated on the ordinate. Plant genotypes are indicated at the top of each bar, as follows: plants expressing wild type *ERECTA* alleles in a Col background were Col4*ER* or Col1*ER* (hatched bars); plants expressing wild type *ERECTA* alleles in a Ld background were *ER* (open bars); plants expressing mutant *erecta* alleles in a Col background were either *er105* or *er2* (stippled boxes); and plants expressing the mutant *er1* allele in a Ld background were *er1* (filled boxes). Columns designated a,b are data from two experiments where plants were grown in soil in the absence of fertiliser. The set of columns at the right of the figure are from a third experiment where the same plants were grown in soil comprising fertiliser. Data indicate that, in a Col background, the *er105* mutation and *er2* mutation enhances stomatal density, which in part accounts for the enhanced stomatal conductances and reduced transpiration efficiencies of plants expressing these alleles (Figures 1b and 1c). The general effects of these alleles is not dependent on the nutrient status of the soil. In contrast, the *er1* allele enhanced stomatal density of Ld plants in only one case when fertiliser was absent, suggesting that enhanced stomatal aperture may account for the enhanced stomatal conductances and reduced transpiration efficiencies of Ld-*er1* plants (Figures 1b, 1c).

Figure 2b is a graphical representation showing the epidermal cell size (surface area, μm^2) for several genotypes of *A. thaliana* in three independent experiments. The genetic backgrounds and genotypes of plants are indicated on the x-axis and at the tops of each column, respectively, as in the legend to Figure 2a. The ordinate indicates epidermal cell size. Columns designated a,b are data from two experiments where plants were grown in soil in the absence of fertiliser. The set of columns at the right of the figure are from a third experiment where the same plants were grown in soil comprising fertiliser. Data indicate that, in a Col background, the *er105* mutation and *er2* mutation significantly reduce epidermal cell size ie increase the number of epidermal cells per unit leaf area. This reveals that the ER gene has effects on leaf histogenesis which, beyond their consequences on stomatal densities, may also directly affect leaf capacity for photosynthesis and therefore transpiration efficiency, (Figures 1b and 1c). The general effects of these alleles is not dependent on the nutrient status of the soil. In contrast, the *er1* allele reduced epidermal cell size of Ld plants in only one case when fertiliser was absent, suggesting that enhanced stomatal aperture accounts for the enhanced stomatal conductances and reduced transpiration efficiencies of Ld-*er1* plants (Figures 1b, 1c).

Figure 2c is a graphical representation showing the stomatal index for several genotypes of *A. thaliana* in three independent experiments. The genetic backgrounds and genotypes of plants are indicated on the x-axis and at the tops of each column, respectively, as in the legend to Figure 2a. The ordinate indicates stomatal index, as determined from the ratio of stomatal density to epidermal cell density. Columns designated a,b are data from two experiments where plants were grown in soil in the absence of fertiliser. The set of columns at the right of the figure are from a third experiment where the same plants were grown in soil comprising fertiliser. Data indicate that the *er* mutations tested do not significantly modify stomatal index, because increases in stomatal density are correlated to increases in epidermal cell numbers in the mutant plants. Accordingly, the *ER* gene does not appear to directly modify stomatal development *per se*.

SUMMARY OF THE INVENTION

In work leading up to the present invention, the inventors sought to elucidate the specific genetic determinants of plant transpiration efficiency. In plants, the

development of molecular genetic markers, such as, for example, genetic markers that map to a region of the genome of a crop plant, such as, for example, a region of the rice genome, maize genome, barley genome, sorghum genome, or wheat genome, or a region of the tomato genome or of any Brassicaceae ,assists in the production of plants having enhanced transpiration efficiency (Edwards *et al.*, *Genetics 116*, 113 - 125, 1987; Paterson *et al.*, *Nature 335*, 721-726, 1988).

The present inventors identified a locus that is linked to the genetic variation in transpiration efficiency in plants. To elucidate a locus associated with the transpiration efficiency of plants, the inventors established experimental conditions and sampling procedures to determine the contribution to total transpiration efficiency of the factors influencing this phenotype, and, more particularly, the genetic contribution to the total variation in transpiration efficiency. Factors influencing transpiration efficiency include, for example, genotype of the plant, environment (eg. temperature, light, humidity, boundary layer around the leaves, root growth conditions), development (eg. age and/or stage and/or posture of plants that modifies gas exchange and/or carbon metabolism), and seed-specific factors (Masle et al. 1993, op. cit). The screens developed by the inventors were also used to survey mutant and wild type populations for variations in transpiration efficiency and to identify ecotypes having contrasting transpiration efficiencies including the parental lines that had been used by Lister and Dean (1993) . The transpiration efficiencies of the members of Lister and Dean's (1993) Recombinant Inbred Line (RIL) mapping population were then determined, and linkage analyses were performed against genetic markers to determine the chromosome regions that are linked to genetic variation in transpiration efficiency, thereby identifying a locus conditioning transpiration efficiency.

In the exemplified embodiment of the invention, there is provided a locus associated with transpiration efficiency of *A. thaliana*, said locus defined as the *ERECTA* locus on *A. thaliana* chromosome 2. The present invention clearly extends to homologs of the *A. thaliana ERECTA* locus from other plant species, identified using the methods described herein.

Accordingly, one aspect of the invention provides a locus associated with the genetic variation in transpiration efficiency of a plant, wherein said locus comprises a nucleotide sequence linked genetically to the *A. thaliana ERECTA* locus.

As used herein, the terms " genetically linked" and "map to" shall be taken to refer to a sufficient genetic proximity between a linked nucleic acid comprising a gene, allele, marker or other nucleotide sequence and nucleic acid comprising all or part of the *A. thaliana ERECTA* locus or all or part of a homolog of the *A. thaliana ERECTA* locus from another plant species, to permit said linked nucleic acid to be useful for determining the presence of a particular allele of said *A. thaliana ERECTA* locus or said homolog. Those skilled in the art will be aware that for such linked nucleic acid to be used in this manner, it must be sufficiently close to said locus not to be in linkage disequilibrium or to have a high recombination frequency between said linked nucleic acid and said locus. Preferably, the linked nucleic acid and the locus are less than about 25cM apart, more preferably less than about 10cM apart, even more preferably less than about 5cM apart, still even more preferably less than about 3cM apart and still even more preferably less than about 1cM apart.

Preferably, all or part of the locus of the invention is provided as recombinant or isolated nucleic acid, such as, for example, in the form of a gene construct (eg. a recombinant plasmid or cosmid), to facilitate germplasm screening.

The *ERECTA* locus or a gene that is linked to the *ERECTA* locus is particularly useful in a breeding program, to predict the transpiration efficiency of a plant, or alternatively, as a selective breeding marker to select plants having enhanced transpiration efficiency. Once mapped, marker-assisted selection (MAS) is used to introduce the *ERECTA* locus or markers linked thereto into a wide variety of populations. MAS has the advantage of reducing the breeding population size required, and the need for continuous recurrent testing of progeny, and the time required to develop a superior line.

Accordingly, a second aspect of the invention provides a method of selecting a plant having enhanced transpiration efficiency, comprising:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant; and

(b) selecting a plant that comprises or expresses a gene that maps to the locus.

Preferably, this aspect of the invention provides a method of selecting a plant having enhanced transpiration efficiency, comprising:

(a) identifying a locus on the Arabidopsis chromosome 2 (46-50.7 cM) associated with genetic variation in transpiration efficiency in a plant; and

(b) selecting a plant that comprises or expresses a gene that maps to the locus.

An alternative embodiment provides a method of selecting a plant having enhanced transpiration efficiency, comprising selecting a plant that comprises or expresses a functionally equivalent homolog of a protein-encoding region of the *ERECTA* gene of *A. thaliana*.

As exemplified herein, the inventors also identified specific genes or alleles that are linked to the *ERECTA* locus and determine the transpiration efficiency of a plant. More particularly, the transpiration efficiencies of near-isogenic lines, wherein each line carries a mutation within the *ERECTA* locus were determined, thereby providing the genetic contribution of genes or alleles at the *ERECTA* locus to transpiration efficiency. This analysis allowed the inventors to assess the genetic contribution of particular alleles to transpiration efficiency, thereby determining allelic variants that are linked to a particular transpiration efficiency. Thus, the elucidation of the *ERECTA* locus for transpiration efficiency in plants facilitated the fine mapping and determination of allelic variants that modulate transpiration efficiency.

Accordingly, a third aspect of the invention provides a method of identifying a gene that determines the transpiration efficiency of a plant comprising:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant;

(b) identifying a gene or allele that is linked to said locus, wherein said gene or allele is a candidate gene or allele for determining the transpiration efficiency of a plant; and

(c) determining the transpiration efficiencies of a panel of plants, wherein not all members of said panel comprise said gene or allele, and wherein variation in transpiration efficiency between the members of said panel indicates that said gene is involved in determining transpiration efficiency.

In another embodiment, the method comprises:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant;

(b) identifying multiple alleles of a gene that is linked to said locus, wherein said gene is a candidate gene involved for determining the transpiration efficiency of a plant; and

(c) determining the transpiration efficiencies of a panel of plants, wherein each member of said panel comprises at least one of said multiple alleles, wherein variation in transpiration efficiency between the members of said panel indicates that said gene is involved in determining transpiration efficiency.

Preferably, the identified gene or allele identified by the method described in the preceding paragraph is an *ERECTA* allele, an *erecta* allele, a homolog of *ERECTA* allele, or a homolog of *erecta* allele, wherein said homolog is from a plant species other than *A. thaliana*.

The identified gene or allele, including any homologs from a plant other than *A. thaliana*, such as, for example, the wild-type *ERECTA* allele or a homolog thereof, is useful for the production of novel plants. Such plants are produced, for example, using recombinant techniques, or traditional plant breeding approaches such as introgression.

Accordingly, a still further aspect of the present invention provides a method of enhancing the transpiration efficiency of a plant comprising ectopically expressing in a plant an isolated gene comprising a nucleotide sequence that is homologous to a protein-encoding region of a gene of *A. thaliana* that maps to the *ERECTA* locus on chromosome 2. A related embodiment of the invention provides a method of enhancing the transpiration efficiency of a plant comprising introgressing into said plant a nucleic acid comprising a nucleotide sequence that is homologous to a

protein-encoding region of a gene of *A. thaliana* that maps to the *ERECTA* locus on chromosome 2.

A further aspect of the invention provides for the use of an isolated gene comprising a nucleotide sequence that is homologous to a protein-encoding region of a gene of *A. thaliana* that maps to the *ERECTA* locus on chromosome 2 in the preparation of a gene construct for enhancing the transpiration efficiency of a plant.

A fifth aspect of the present invention provides a plant having enhanced transpiration efficiency, wherein said plant is produced by a method described herein.

DETAILED DESCRIPTION OF THE INVENTION

Loci for transpiration efficiency and their identification

One aspect of the invention provides a locus associated with the genetic variation in transpiration efficiency of a plant, wherein said locus comprises a nucleotide sequence linked genetically to the *ERECTA* locus on chromosome 2 or a homolog thereof.

As used herein, the term "locus" shall be taken to mean the location of one or more genes in the genome of a plant that affects a quantitative characteristic of the plant, in particular the transpiration efficiency of a plant. In the present context, a "quantitative characteristic" is a phenotype of the plant for which the phenotypic variation among different genotypes is continuous and cannot be separated into discrete classes, irrespective of the number of genes that determine or control the phenotype, or the magnitude of genetic effects that single gene has in determining the phenotype, or the magnitude of genetic effects of interacting genes.

By "associated with the genetic variation in transpiration efficiency of a plant" means that a locus comprises one or more genes that are expressed to determine or regulate the transpiration efficiency of a plant, irrespective of the actual rate of transpiration achieved by the plant under a specified environmental condition.

The present invention clearly contemplates the presence of multiple genes that are genetically linked or map to the specified *ERECTA* locus on chromosome 2. Without being bound by any theory or mode of action, such multiple linked genes may interact, such as, for example, by epistatic interaction, to determine the transpiration efficiency phenotype.

The present invention also contemplates the presence of different alleles of any gene that is linked to the *ERECTA* locus, wherein said allele is expressed to determine the transpiration efficiency phenotype. In one embodiment, such alleles are identified by detecting a particular transpiration efficiency phenotype that is linked to the expression of the particular allele. Alternatively, or in addition, the different alleles linked to a locus are identified by detecting a structural polymorphism in DNA (eg. a restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), single strand chain polymorphism (SSCP), and the like), that is linked to a particular transpiration efficiency phenotype.

The present invention clearly encompasses all interacting genes and/or alleles that are genetically linked to an *ERECTA* locus and are expressed to determine a transpiration efficiency phenotype. Such linked interacting genes and/or alleles will map to a region of the genome of a plant that is homologous to a region of the *Arabidopsis thaliana* genome that is associated with the transpiration efficiency of that plant and maps to the *ERECTA* locus. Preferably, such interacting genes and/or alleles comprise a protein-encoding portion of a gene of *A. thaliana* positioned within the *ERECTA* locus of the *Arabidopsis thaliana* genome that is associated with the transpiration efficiency of that plant, or a homologous protein-encoding region from another plant species.

The terms "region of the *Arabidopsis thaliana* genome that is associated with the transpiration efficiency", and "locus of *A. thaliana* that determines the transpiration efficiency" shall be taken to mean that portion of chromosome 2 of *A. thaliana* associated with transpiration efficiency, preferably the region from about 46cM to about 50.7 cM.

Even more preferably, the locus of the invention is linked to or comprises the *ERECTA* allele or the *erecta* alleles, or a protein-encoding portion thereof or a homologous gene from another plant species.

As used herein, the term "*ERECTA*" shall be taken to refer to a wild type *A. thaliana ERECTA* allele or a homolog thereof from another plant species.

For the purposes of nomenclature, the nucleotide sequence of the *Arabidopsis thaliana ERECTA* protein-encoding region, including 5'-untranslated region (UTR) and 3'-UTR, is provided herein as SEQ ID NO: 1. The amino acid sequence of the polypeptide encoded by SEQ ID NO: 1 is set forth herein as SEQ ID NO: 2.

The term "*erecta*" shall be taken to mean any allelic variant of the wild-type *ERECTA* allele that modifies transpiration efficiency of *A. thaliana* or a homolog thereof from another plant species. Preferred *erecta* alleles contemplated herein include an *A. thaliana erecta* allele selected from the group consisting of: *er*, *er1*, *er2* allele, *er101* allele, *er102*allele, *er103* allele, *er104* allele, *er105* allele, *er2* allele, *er108* allele, *er109* allele, *er110* allele, *er111* allele, *er112* allele, *er113* allele, *er114* allele. *er115* allele, *er116* allele, *er117* allele, *er118* allele, *er119* allele, *er120* allele, *er121* allele, *er122* allele, *er123* allele (Lease *et al. New Phytologist 151*, 133-143, 2001) and a homolog of any one of said alleles.

Those skilled in the art are aware that the terms "homolog" and "ortholog" refer to functional equivalent units. In the present context, a homolog or ortholog of a gene that maps to the *ERECTA* locus shall be taken to mean any gene from a plant species other than *A. thaliana* that is functionally equivalent to a gene that maps to the exemplified *A. thaliana ERECTA* locus, and comprises a protein-encoding region in its native plant genome that shares a degree of structural identity or similarity with a protein-encoding region that is linked to the *A. thaliana ERECTA* locus.

Preferably, a homologous or orthologous gene from a plant other than *A. thaliana* will be associated with the transpiration efficiency of said plant and be linked to a protein-encoding region in its native plant genome that comprises a nucleotide sequence having at least about 55% overall sequence identity to a protein-encoding

region linked to the *ERECTA* locus. Even more preferably, the percentage identity will be at least about 59-61% or 70% or 80%, still more preferably at least about 90%, and even still more preferably at least about 95%.

In determining whether or not two nucleotide sequences fall within a particular percentage identity limitation recited herein, those skilled in the art will be aware that it is necessary to conduct a side-by-side comparison or multiple alignment of sequences. In such comparisons or alignments, differences may arise in the positioning of non-identical residues, depending upon the algorithm used to perform the alignment. In the present context, reference to a percentage identity between two or more nucleotide sequences shall be taken to refer to the number of identical residues between said sequences as determined using any standard algorithm known to those skilled in the art. For example, nucleotide sequences may be aligned and their identity calculated using the BESTFIT program or other appropriate program of the Computer Genetics Group, Inc., University Research Park, Madison, Wisconsin, United States of America (Devereaux *et al, Nucl. Acids Res. 12*, 387-395, 1984).

Alternatively, or in addition, a homologous or orthologous gene from a plant other than *A. thaliana* will be associated with the transpiration efficiency of said plant and be linked to a protein-encoding region in its native plant genome that comprises a nucleotide sequence that encodes a polypeptide having at least about 55% overall sequence identity to a polypeptide encoded by a protein-encoding region linked to the *ERECTA* locus. Preferably, the percentage identity at the amino acid level will be at least about 59-61% or 70% or 80%, more preferably at least about 90%, and still more preferably at least about 95%.

In determining whether or not two amino acid sequences fall within these percentage limits, those skilled in the art will be aware that it is necessary to conduct a side-by-side comparison or multiple alignment of sequences. In such comparisons or alignments, differences will arise in the positioning of non-identical residues, depending upon the algorithm used to perform the alignment. In the present context, reference to a percentage identity or similarity between two or more amino acid sequences shall be taken to refer to the number of identical and similar residues respectively, between said sequences as determined using any standard algorithm

known to those skilled in the art. For example, amino acid sequence identities or similarities may be calculated using the GAP program and/or aligned using the PILEUP program of the Computer Genetics Group, Inc., University Research Park, Madison, Wisconsin, United States of America (Devereaux *et al*, 1984, *supra*). The GAP program utilizes the algorithm of Needleman and Wunsch, *J. Mol. Biol. 48*, 443-453, 1970, to maximize the number of identical/similar residues and to minimize the number and length of sequence gaps in the alignment. Alternatively or in addition, wherein more than two amino acid sequences are being compared, the ClustalW program of Thompson *et al., Nucl. Acids Res. 22*, 4673-4680, 1994, is used.

Alternatively, or in addition, a homologous or orthologous gene from a plant other than *A. thaliana* will be associated with the transpiration efficiency of said plant and be linked to a protein-encoding region in its native plant genome that hybridizes to nucleic acid that comprises a sequence complementary to a protein-encoding region linked to the *A. thaliana ERECTA* locus. Preferably, such homologs or orthologs will be identified by hybridization under at least low stringency conditions, and more preferably under at least moderate stringency or high stringency hybridization conditions.

For the purposes of defining the level of stringency, a low stringency is defined herein as being a hybridization or a wash carried out in 6xSSC buffer, 0.1% (w/v) SDS at 28°C or alternatively, as exemplified herein. Generally, the stringency is increased by reducing the concentration of salt in the hybridization or wash buffer, such as, for example, by reducing the concentration of SSC. Alternatively, or in addition, the stringency is increased, by increasing the concentration of detergent (eg. SDS). Alternatively, or in addition, the stringency is increased, by increasing the temperature of the hybridization or wash. For example, a moderate stringency can be performed using 0.2xSSC to 2xSSC buffer, 0.1% (w/v) SDS, at a temperature of about 42°C to about 65°C. Similarly, a high stringency can be performed using 0.1xSSC to 0.2xSSC buffer, 0.1% (w/v) SDS, at a temperature of at least 55°C. Conditions for performing nucleic acid hybridization reactions, and subsequent membrane washing, are well understood by one normally skilled in the art. For the purposes of further clarification only, reference to the parameters affecting hybridization between nucleic acid molecules is found in Ausubel *et al., In:* Current

Protocols in Molecular Biology, Greene/Wiley, New York USA, 1992, which is herein incorporated by reference.

Preferably, the homologous gene is derived from, or present in, the genome of a plant that is desiccation or drought intolerant, or poorly adapted for growth in dry or arid environments, or that suffers from reduced vigor or growth during periods of reduced rainfall or drought, or from the genome of a plant with increased growth rate or growth duration or partitioning of C to shoot and harvested parts under well-watered conditions.

More preferably, the homologous gene is derived from, or present in, the genome of a brassica plant, broad acre crop plant, perennial grass (eg. of the subfamily Pooidaea, or the Tribe Poeae), or tree. Even more preferably, the homologous locus is in the genome of a plant selected from the group consisting of barley, wheat, rye, sorghum, rice, maize, *Phalaris aquatica*, *Dactylus glomerata*, *Lolium perenne*, *Festuca arundinacea*, cotton, tomato, soybean, oilseed rape, poplar, and pine.

A particularly preferred homolog of a gene that maps to the exemplified *Arabidopsis thaliana ERECTA* locus is derived from chromosome 6 of rice (*Oryza sativa*), and, more preferably is linked to an *ERECTA* gene derived from rice. For the purposes of nomenclature, the protein-encoding region of the rice *ERECTA* gene is provided herein as SEQ ID NO: 3. The amino acid sequence of the polypeptide encoded by SEQ ID NO: 3 is set forth herein as SEQ ID NO: 4.

Another particularly preferred homolog of a gene that maps to the exemplified *Arabidopsis thaliana ERECTA* locus is derived from the genome of *Sorghum bicolor*, and, more preferably is linked to an *ERECTA* gene derived from sorghum. For the purposes of nomenclature, the protein-encoding region of the sorghum *ERECTA* gene is provided herein as SEQ ID NO: 5. The amino acid sequence of the polypeptide encoded by SEQ ID NO: 5 is set forth herein as SEQ ID NO: 6.

Another particularly preferred homolog of a gene that maps to the exemplified *Arabidopsis thaliana ERECTA* locus is derived from the genome of *A. thaliana*, and, more preferably is linked to an *ERECTA* gene derived from *A. thaliana*. For the

purposes of nomenclature, the protein-encoding region of the *A. thaliana ERECTA* homologue is provided herein as SEQ ID NO: 7. The amino acid sequence of the polypeptide encoded by SEQ ID NO: 7 is set forth herein as SEQ ID NO: 8.

Another particularly preferred homolog of an *ERECTA* gene is provided herein as SEQ ID NO: 9. The amino acid sequence of the polypeptide encoded by SEQ ID NO: 9 is set forth herein as SEQ ID NO: 10.

A number of mapping methods for determining useful loci and estimating their effects have been described (eg. Edwards *et al.*, *Genetics 116*, 113-125, 1987; Haley and Knott, *Heredity 69*, 315-324, 1992; Jiang and Zeng, *Genetics 140*, 1111-1127, 1995; Lander and Botstein, *Genetics 121*, 185-199, 1989; Jansen and Stam, *Genetics 136*, 1447-1455, 1994; Utz and Melchinger, *In:* Biometrics in Plant Breeding: Applications of Molecular Markers. Proc. Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, 6 - 8 July 1994, Wageningen, The Netherlands, (J.W. van Ooijen and J. Jansen, eds), pp195-204, 1994; Zeng, *Genetics 136*, 1457-1468, 1994). In the present context, these methods are applied to identify the major component(s)of the total genetic variance that contribute(s) to the variation in transpiration efficiency of a plant, such as, for example, determined by the measurement of carbon isotope discrimination (Δ). More particularly, the segregation of known markers is used to map and/or characterize an underlying locus associated with transpiration efficiency. The locus method involves searching for associations between the segregating molecular markers and transpiration efficiency in a segregating population of plants, to identify the linkage of the marker to the locus.

To discover a marker/locus linkage, a segregating population is required. Experimental populations, such as, for example, an F2 generation, a backcross (BC) population, recombinant inbred line (RIL), or double haploid line (DHL), can be used as a mapping population. Bulk segregant analysis, for the rapid detection of markers at specific genomic regions using segregating populations, is described by Michelmoore *et al.*, *Proc. Natl Acad. Sci. (USA) 88*, 9828-9832, 1991. In the case of F2 mapping populations, F2 plants are used to determine genotype, and F2 families to determine phenotype. Recombinant inbred lines are produced by single-seed

descent. Recombinant inbred lines, such as, for example, the F9 RILs of *A. thaliana* (eg. Lister and Dean, *Plant J., 4*, 745-750, 1993) will be known to those skilled in the art. Near isogenic lines (NILs) are used for fine mapping, and to determine the effect of a particular locus on transpiration efficiency. An advantage of recombinant inbred lines and double haploid lines is that they are permanent populations, and as a consequence, provide for replication of the contribution of a particular locus to the transpiration efficiency phenotype.

As for statistical methods, Single Marker Analysis (Point Analysis) is used to detect a locus in the vicinity of a single genetic marker. The mean transpiration efficiencies of a population of plants segregating for a particular marker, are compared according to the marker class. The difference between two mean transpiration efficiencies provides an estimate of the phenotypic effect of substituting one allele for another allele at the locus. To determine whether or not the inferred phenotypic effect is significantly different from zero, a simple statistical test, such as t-test or F-test, is used. A significant value indicates that a locus is located in the vicinity of the marker. Single point analysis does not require a complete molecular linkage map. The further locus is from the marker, the less likely it is to be detected statistically, as a consequence of recombination between the marker and the gene.

In the Anova, t-test or GLM approach, the association between marker genotype and transpiration efficiency phenotype comprises:

(i) classifying progeny of a segregating population of plants by marker genotype, such as for example, using RFLP, AFLP, SSCP, or microsatellite analyses, thereby establishing classes of plants;

(ii) comparing the mean transpiration efficiencies of classes of plants in the segregating population, using a t-test, GLM or ANOVA; and

(iii) determining the significance of the differences in the mean at (ii), wherein a significant difference indicates that the marker is linked to the locus for transpiration efficiency.

As will be known to those skilled in the art, the difference between the means of the classes provides an estimate of the effect of the locus in determining the transpiration efficiency of a class.

In the regression approach, the association between marker genotype and phenotype is determined by a process comprising:

(i) assigning numeric codes to marker genotypes; and

(ii) determining the regression value r for transpiration efficiency against the codes, wherein a significant value for r indicates that the marker is linked to the locus for transpiration efficiency, and wherein the regression slope estimate of the effect of a particular locus on transpiration efficiency.

For QTL interval mapping, the Mapmaker algorithm developed by Lincoln *et al.*, Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: A tutorial and reference manual. Whitehead Institute for Biomedical Research, Cambridge, MA, USA, 1993, can be used. The principle behind interval mapping is to test a model for the presence of a QTLat many positions between two mapped marker loci. This model is a fit of a presumptive QTL to transpiration efficiency, wherein the suitability of the fit is tested by determining the maximum likelihood that a QTL for transpiration efficiency lies between two segregating markers. For example, in the case of a QTL located between two segregating markers, the 2-loci marker genotypes of segregating progeny will each contain mixtures of QTL genotypes. Accordingly, it is possible to search for loci parameters that best approximate the distribution in transpiration efficiency for each marker class. Models are evaluated by computing the likelihood of the observed distributions with and without fitting a QTL effect. The map position of a QTL is determined as the maximum likelihood from the distribution of likelihood values (LOD scores: ratio of likelihood that the effect occurs by linkage: likelihood that the effect occurs by chance), calculated for each locus.

Interval mapping by regression (Haley and Knott., *Heredity 69*, 315-324, 1992) is a simplification of the maximum likelihood method *supra* wherein basic QTL analysis or regression on coded marker genotypes is performed, except that phenotypes are regressed on the probability of a QTL genotype as determined from the linkage between transpiration efficiency and the nearest flanking markers. In most cases, regression mapping gives estimates of QTL position and effect that are almost

identical to those given by the maximum likelihood method. The approximation deviates only at places where there are large gaps, or many missing genotypes.

In the composite interval mapping (CIM) method (Jansen and Stam, *Genetics 136*, 1447-1455, 1994; Utz and Melchinger, 1994, *supra*; Zeng, *Genetics 136*, 1457-1468, 1994), the analysis is performed in the usual way, except that the variance from other QTLs are accounted for by including partial regression more power and precision than simple interval mapping, because the effects of other QTls are not present as residual variance. CIM can remove the bias that can be caused by the QTLs that are linked to the position being tested.

Publicly available software are used to map a locus for transpiration efficiency. Such software include, for example, the following:

(i) MapMaker/QTL (ftp://genome.wi.mit.edu/pub/mapmaker3/), for analyzing F2 or backcross data using standard interval mapping;

(ii) MQTL, for composite interval mapping in multiple environments or for performing simple interval mapping using homozygous progeny (eg. double haploids, or recombinant inbred lines);

(iii) PLABQTL (Utz and Melchinger, PLABlocus Version 1.0. A computer program to map QTL, Institut für Pflanzenzüchtung, Saatgutforschung und Populationsgenetik, Universität Hohenheim, 70593 Stuttgart, Germany, 1995; http://www.uni-hohenheim.de/~ipspwww/soft.html) for composite interval mapping and simple interval mapping of a locus in mapping populations derived from a bi-parental cross by selfing, or in double haploids;

(iv) QTL Cartographer (http://statgen.mcsu.edu/qtlcart/cartographer.html) for single-marker regression, interval mapping, or composite interval mapping, using F2 or backcross populations;

(v) MapQTL (http://www.cpro.dlo.nl/cbw/); Qgene for performing either single-marker regression or interval regression to map loci; and

(vi) SAS for detecting a locus by identifying associations between marker genotype and transpiration efficiency by a single marker analysis approach such as ANOVA, t-test, GLM or REG.

In a particularly preferred embodiment, QTL cartographer or MQTL is used to identify a locus associated with the transpiration efficiency of plants.

Those skilled in the art will also be aware that it is possible to detect multiple interacting alleles or genes for a particular trait, such as, for example, using composite interval mapping approaches. To achieve this end, the composite interval mapping may be repeated to look for additional loci. Alternatively, or in addition, two or more distinct regions of the genome can be nominated as candidate loci, and a gamete relationship matrix constructed for each candidate locus, and a 2-locus regression performed for each pair of loci, determining a best fit for the interacting effects between the two loci or aleles at those loci, including any dominance or additive effects. The algorithm described by Carlborg *et al.*, *Genetics* (2000) can be used for simultaneous mapping. In the present context, such an analysis is performed with reference to the segregation of transpiration efficiency phenotypes in the segregating population.

Use of the ERECTA locus to enhance transpiration efficiency of plants

As will be known to those skilled in the art, a single locus, if present in the genome of a plant, can have a significant influence on the phenotype of the plant. For example, Grandillo *et al.*, *Theor. Appl. Genet. 99*, 978-987, 1999, showed that for tomato a selection made from a total 28 loci determining fruit size and weight explained 20% of the total phenotypic variance in this trait.

Accordingly, a second aspect of the invention provides a method of selecting a plant having enhanced transpiration efficiency, comprising:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant; and

(b) selecting a plant that comprises or expresses a gene that maps to the locus.

By "enhanced transpiration efficiency" is meant that the plant loses less water per unit of dry matter produced, or alternatively, produces an enhanced amount of dry matter per unit of water transpired, relative to a counterpart plant. By "counterpart plant" is meant a plant having a similar or near-identical genetic background, such as, for example, a near-isogenic plant, a sibling, or parent.

In accordance with this aspect of the invention, a locus is identified by conventional locus mapping means, and/or by homology searching for genes that map to the *ERECTA* locus on chromosome 2 of the *A. thaliana* genome, such as, for example, by searching for homologs of the *A. thaliana ERECTA* allele or *erecta* allele as described herein above.

Preferably, to select a plant that comprises or expresses the appropriate gene, marker-assisted selection (MAS) is used. As will be known to those skilled in the art, once a particular locus has been identified, genetic or physical markers that are linked to the locus can be readily identified and used to confirm the presence of the locus in breeding populations. For a locus that is flanked by two tightly-linked markers that recombine only at a low frequency, the presence of the flanking markers is indicative of the presence of the locus.

For marker-assisted selection, it is preferred that the marker is a genetic marker (eg. a gene or allele), or a physical marker (eg. leaf hairiness or pod shape), or a molecular marker such as, for example, a restriction fragment length polymorphism (RFLP), a restriction (RAPD), amplified fragment length polymorphism (AFLP), or a short sequence repeat (SSR) such as a microsatellite, or SNP. It is also within the scope of the invention to utilize any hybridization probe or amplification primer comprising at least about 10 nucleotides in length derived from a chromosome region that is linked in the genome of a plant to the *ERECTA* locus, as a marker to select plants. Those skilled in the art will readily be able to determine such probes or primers based upon the disclosure herein, particularly for those plant genomes which may have sufficient chromosome sequence in the region of interest in the genome (eg. *A. thaliana*, rice, cotton, barley, wheat, sorghum, maize, tomato, etc).

For flanking markers that are not tightly linked, such that there is a large recombination distance there between, the presence of the appropriate gene is assessed by identifying those plants having both flanking markers and then selecting from those plants a plant having an enhanced transpiration efficiency. Naturally, the greater the distance between two markers, the larger the population of plants required to identify a plant having both markers, the intervening locus and a gene

within said locus. Those skilled in the art will readily be able to determine the population size required to identify a plant having a particular transpiration efficiency, based upon the recombination units (cM) between two markers.

Transpiration efficiency is determined by any means known to the skilled artisan. Preferably, transpiration efficiency is determined by measuring dry matter accumulation in the plant by gravimetric means, or by measuring water loss, or the ratio of CO_2 assimilation rate to stomatal conductance.

In a particularly preferred embodiment, the transpiration efficiency is determined directly, by measuring the ratio of carbon fixed carbon assimilation rate) to water loss (transpiration rate).

In an alternative embodiment, transpiration efficiency is determined indirectly from the carbon isotope discrimination value (Δ). Farquhar *et al., Aust. J. Plant Physiol.* 9,121-137, 1982, showed that carbon isotope discrimination (Δ; a measure of the extent to which the $^{13}C/^{12}C$ ratio of organic matter is less than that of CO_2 in the source air), is an effective indirect measure of transpiration efficiency. The determination of transpiration efficiency in this manner is based upon the constancy of the atmospheric ^{13}C: ^{12}C ratio (about 98.19: 1.11) and the finding that, in C_3 plants at least, ribulose bisphosphate carboxylase (Rubisco) enzymes discriminate against the use of ^{13}C. Thus, $^{13}CO_2$ is less efficiently assimilated than $^{12}CO_2$, and diffuses lessthrough stomata in and out of the leaf. However, when the stomata become nearly closed, the diffusion of $^{13}CO_2$ is more difficult to achieve and, at higher intracellular concentrations of $^{13}CO_2$, this isotope is incorporated into 3-phosphoglycerate, and subsequently into dry matter. As a consequence, carbon isotope discrimination (Δ) is greatest when the overall CO_2 assimilation rate during photosynthesis (A) is small, and stomatal conductance (g_w) to water vapor is large. This relationship is represented by the following algorithm:

$$\Delta\ (^{o}/oo) = 27 - 36A/(G_w \times C_a)$$

wherein C_a is the ambient CO_2 concentration (ie. $[^{12}CO_2 + ^{13}CO_2]$).

For a C_3 plant that exhibits a value in the range of about 4.5 $^{o}/oo$ to about 6.7 $^{o}/oo$ for the term $36A/(G_w \times C_a)$, a 1 $^{o}/oo$ change in carbon isotope discrimination (Δ)

corresponds to a change in transpiration efficiency in the range of about 22% to about 15%, respectively.

The negative relationship between carbon isotope discrimination (Δ) and transpiration efficiency has been established for many plant species, including wheat (Farquhar and Richards, *Aust. J. Plant Physiol. 11*, 539-552, 1984; Farquhar *et al., Ann. Rev. Plant Physiol. 40*,388-397, 1989), *Stylosanthes* (Thumma *et al., Proc. 9th Aust. Agronomy Conf.*, Wagga Wagga New South Wales, Australia, 1998), cotton, barley, and rice. Accordingly, a lower carbon isotope discrimination (Δ) value for a test plant relative to a counterpart plant is indicative of enhanced transpiration efficiency.

Alternatively, or in addition, transpiration efficiency is determined by another indicator, such as, for example, leaf temperature, ash content, mineral content, or specific leaf weight (dry matter per unit leaf area). For example, specific leaf weight is positively correlated with transpiration efficiency in peanuts and other species (Virgona *et al., Aust. J. Plant Physiol., 17*, 207-214, 1990; Wright *et al., Crop Sci 34*, 92-97, 1994). Accordingly, a higher specific leaf weight or higher carbon gain rate for a test plant relative to a counterpart plant is indicative of enhanced transpiration efficiency of the test plant.

The presence of the locus can be established by hybridizing a probe or primer that is linked to an *ERECTA* locus, such as, for example, a probe or primer that hybridizes to the identified chromosome 2 region of *A. thaliana* or the identified chromosome 6 region of rice.

Preferably, the presence of the locus is established by hybridizing a probe or primer derived from any one or more of SEQ ID Nos: 1, 3, or 5, or from a homologous gene in another plant, or a complementary sequence to such a sequence, to genomic DNA from the plant, and detecting the hybridization using a detection means.

In one embodiment, detection of the hybridization is performed preferably by labelling a probe with a reporter molecule capable of producing an identifiable signal, prior to hybridization, and then detecting the signal after hybridization.

Preferred reporter molecules include radioactively-labelled nucleotide triphosphates and biotinylated molecules. Preferably, variants of the genes exemplified herein, including genomic equivalents, are isolated by hybridisation under moderate stringency or more preferably, under high stringency conditions, to the probe.

Alternatively, or in addition, hybridization may be detected using any format of the polymerase chain reaction (PCR), including AFLP. For PCR, two non-complementary nucleic acid primer molecules comprising at least about 20 nucleotides in length, and more preferably at least 30 nucleotides in length are hybridized to different strands of a nucleic acid template molecule, and specific nucleic acid molecule copies of the template are amplified enzymatically. Several formats of PCR are described in McPherson *et al.*, *In: PCR A Practical Approach.*, IRL Press, Oxford University Press, Oxford, United Kingdom, 1991, which is incorporated herein by reference.

For enhancing the transpiration efficiency of a plant wherein the locus is polymorphic, such as, for example, an allele, the method *supra* is modified to include the detection of the specific allele(s) linked to the desired enhancement. According to this embodiment, there is provided a method of selecting a plant having enhanced transpiration efficiency, comprising:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant;

(b) identifying a polymorphic marker within said locus that is linked to enhanced transpiration efficiency; and

(c) selecting a plant that comprises or expresses the marker.

Standard means known to the skilled artisan are used to identify a marker within the locus that is linked to enhanced transpiration efficiency. A population of plants that is segregating for the polymorphic marker is generally used, wherein the transpiration efficiency phenotype of plants is then correlated or associated with the presence of a particular allelic form of the marker. As exemplified herein, near-isogenic or recombinant inbred lines of plants were screened to segregate alleles at the *ERECTA* locus and to correlate enhanced transpiration efficiency with the presence of the *ERECTA* allele as opposed to an *erecta* allele.

Suitable markers include any one or more of the markers described herein to be suitable for MAS.

Preferably, the selection of plants in accordance with these embodiment includes the additional step of introducing the locus or polymorphic marker to a plant, such as, for example, by standard breeding approaches or by recombinant means. This may be carried out at the same time, or before, selecting the locus or polymorphic marker.

Recombinant means generally include introducing a gene construct comprising the locus or marker into a plant cell, selecting transformed tissue and regenerating a whole plant from the transformed tissue explant. Means for introducing recombinant DNA into plant tissue or cells include, but are not limited to, transformation using $CaCl_2$ and variations thereof, in particular the method described by Hanahan (1983), direct DNA uptake into protoplasts (Krens *et al, Nature 296,* 72-74, 1982; Paszkowski *et al., EMBO J. 3,* 2717-2722, 1984), PEG-mediated uptake to protoplasts (Armstrong *et al., Plant Cell Rep. 9,* 335-339, 1990) microparticle bombardment, electroporation (Fromm *et al., Proc. Natl. Acad. Sci. (USA), 82,* 5824-5828, 1985), microinjection of DNA (Crossway *et al., Mol. Gen. Genet. 202,* 179-185, 1986), microparticle bombardment of tissue explants or cells (Christou *et al, Plant Physiol. 87,* 671-674, 1988; Sanford, *Part. Sci.Technol. 5,* 27-37, 1988), vacuum-infiltration of tissue with nucleic acid, or in the case of plants, T-DNA-mediated transfer from *Agrobacterium* to the plant tissue as described essentially by An *et al., EMBO J. 4,* 277-284, 1985; Herrera-Estrella *et al.,* Herrera-Estella *et al., Nature 303,* 209-213, 1983; Herrera-Estella *et al., EMBO J. 2,* 987-995, 1983; or Herrera-Estella *et al., In:* Plant Genetic Engineering, Cambridge University Press, N.Y., pp 63-93, 1985..

For microparticle bombardment of cells, a microparticle is propelled into a cell to produce a transformed cell. Any suitable ballistic cell transformation methodology and apparatus can be used in performing the present invention. Exemplary apparatus and procedures are disclosed by Stomp *et al.* (U.S. Patent No. 5,122,466) and Sanford and Wolf (U.S. Patent No. 4,945,050). When using ballistic transformation procedures, the gene construct may incorporate a plasmid capable of replicating in the cell to be transformed.

Examples of microparticles suitable for use in such systems include 1 to 5 micron gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

A whole plant may be regenerated from the transformed or transfected cell, in accordance with procedures well known in the art. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a gene construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (eg., apical meristem, axillary buds, and root meristems), and induced meristem tissue (eg., cotyledon meristem and hypocotyl meristem).

The term "organogenesis", as used herein, means a process by which shoots and roots are developed sequentially from meristematic centres.

The term "embryogenesis", as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed to give homozygous second generation (or T2) transformant, and the T2 plants further propagated through classical breeding techniques.

The generated transformed organisms contemplated herein may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed

cells; clonal transformants (eg., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (eg., in plants, a transformed root stock grafted to an untransformed scion).

Alternatively, the transformed plants are produced by an *in planta* transformation method using *Agrobacterium tumefaciens*, such as, for example, the method described by Bechtold *et al., CR Acad. Sci. (Paris, Sciences de la vie/ Life Sciences) 316*, 1194-1199, 1993 or Clough *et al., Plant J 16:* 735-74, 1998, wherein *A. tumefaciens* is applied to the outside of the developing flower bud and the binary vector DNA is then introduced to the developing microspore and/or macrospore and/or the developing seed, so as to produce a transformed seed. Those skilled in the art will be aware that the selection of tissue for use in such a procedure may vary, however it is preferable generally to use plant material at the zygote formation stage for *in planta* transformation procedures.

Identification of genes for determing the transpiration efficiency of a plant

As exemplified herein, the inventors also identified specific genes or alleles that are linked to the *ERECTA* locus and determine the transpiration efficiency of a plant. More particularly, the transpiration efficiencies of near-isogenic lines, wherein each line carries a mutation within a target locus in the region of a locus associated with transpiration efficiency, were determined, thereby providing the genetic contribution of that locus to transpiration efficiency. This analysis allowed the inventors to assess the genetic contribution of particular alleles to transpiration efficiency, thereby determining allelic variants that are linked to a particular transpiration efficiency. Thus, the elucidation of the *ERECTA* locus for transpiration efficiency in plants facilitated the fine mapping and determination of allelic variants that determine transpiration efficiency.

Accordingly, a third aspect of the invention provides a method of identifying a gene that determines the transpiration efficiency of a plant.

In one embodiment, the method comprises:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant;

(b) identifying a gene or allele that is linked to said locus, wherein said gene or allele is a candidate gene or allele for determining the transpiration efficiency of a plant; and

(c) determining the transpiration efficiencies of a panel of near isogenic plants, wherein not all members of said panel comprise said gene or allele, and wherein variation in transpiration efficiency between the members of said panel indicates that said gene is involved in determining transpiration efficiency.

In another embodiment, the method comprises:

(a) identifying a locus associated with genetic variation in transpiration efficiency in a plant;

(b) identifying multiple alleles of a gene that is linked to said locus, wherein said gene is a candidate gene involved for determining the transpiration efficiency of a plant; and

(c) determining the transpiration efficiencies of a panel of near isogenic plants, wherein each member of said panel comprises at least one of said multiple alleles, wherein variation in transpiration efficiency between the members of said panel indicates that said gene is involved in determining transpiration efficiency.

In the present context, the term "near isogenic plants" shall be taken to mean a population of plants having identity over a substantial proportion of their genomes, notwithstanding the presence of sufficiently few differences to permit the contribution of a distinct allele or gene to the transpiration efficiency of a plant to be determined by a comparison of the transpiration efficiency phenotypes of the population. As will be known to the skilled artisan, recombinant inbred lines, lines produced by introgression of a gene followed by several generations of backcrossing, or siblings, are suitable near-isogenic lines for the present purpose.

Preferably, the identified gene or allele identified by the method described in the preceding paragraph is selected from the group consisting of *ERECTA* gene, *Erecta*

alleles, homologs of *ERECTA*, , wherein said homologs are from plants species other than *A. thaliana*.

In a particularly preferred embodiment, the identified gene or allele will comprise a nucleotide sequence selected from the group consisting of:

(a) a sequence having at least about 55% identity to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9;

(b) a sequence encoding an amino acid sequence having at least about 55% identity to an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10; and

(c) a sequence complementary to (a) or (b).

Preferably, the percentage identity is at least about 59-61% or 70% or 80%, more preferably at least about 90%, and even more preferably at least about 95% or 99%. In a particularly preferred embodiment, the identified gene or allele comprises a nucleotide sequence selected from the group consisting of:

(a) a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9;

(b) a sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10; and

(c) a sequence complementary to (a) or (b).

Enhancement of transpiration efficiency using isolated genes

The identified gene or alleles, including any homologs from a plant other than *A. thaliana*, such as, for example, the wild-type *ERECTA* allele, or a homolog thereof, is useful for the production of novel plants. Such plants are produced, for example, using recombinant techniques, or traditional plant breeding approaches such as by introgression.

Accordingly, a fourth aspect of the present invention provides a method of enhancing the transpiration efficiency of a plant comprising ectopically expressing

in a plant an isolated gene comprising a nucleotide sequence that is homologous to a protein-encoding region of a gene that is linked to the *A. thaliana ERECTA* locus on chromosome 2.

In a particularly preferred embodiment, the isolated gene comprises a nucleotide sequence selected from the group consisting of:

(a) a sequence having at least about 55% identity to a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9;

(b) a sequence encoding an amino acid sequence having at least about 55% identity to an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10; and

(c) a sequence complementary to (a) or (b).

Preferably, the percentage identity is at least about 59-61% or 70% or 80%, more preferably at least about 90%, and even more preferably at least about 95% or 99%.

In a particularly preferred embodiment, the isolated gene or allele comprises a nucleotide sequence selected from the group consisting of:

(a) a sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9;

(b) a sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10; and

(c) a sequence complementary to (a) or (b).

To ectopically express the isolated gene in a plant, the protein-encoding portion of the gene is generally placed in operable connection with a promoter sequence that is operable in the plant, which may be the endogenous promoter or alternatively, a heterologous promoter, and a transcription termination sequence, which also may be the endogenous or an heterologous sequence relative to the gene of interest. The promoter and protein-encoding portion and transcription termination sequence are generally provided in the form of a gene construct, to facilitate introduction and

maintenance of the gene in a plant where it is to be ectopically expressed. Numerous vectors suitable for introducing genes into plants have been described and are readily available. These may be adapted for expressing an isolated gene in a plant to enhance transpiration efficiency therein.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical eukaryotic genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional regulatory elements (ie. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. In the present context, the term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of said sense molecule in a cell. Preferred promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression and/or to alter the spatial expression and/or temporal expression of a nucleic acid molecule to which it is operably connected. For example, copper-responsive regulatory elements may be placed adjacent to a heterologous promoter sequence driving expression of a nucleic acid molecule to confer copper inducible expression thereon.

Placing a nucleic acid molecule under the regulatory control of a promoter sequence means positioning said molecule such that expression is controlled by the promoter sequence. A promoter is usually, but not necessarily, positioned upstream or 5' of the protein-encoding portion of the gene that it regulates. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2 kb of the start site of transcription of the structural protein-encoding nucleotide sequences, or a chimeric gene comprising same. In the construction of heterologous promoter/structural gene combinations it is generally preferred to position the promoter at a distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting, ie., the gene from which the promoter is derived. As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the

positioning of the element in its natural setting, ie., the genes from which it is derived. Again, as is known in the art, some variation in this distance can also occur.

Promoters suitable for use in gene constructs of the present invention include those promoters derived from the genes of viruses, yeasts, moulds, bacteria, insects, birds, mammals and plants which are capable of functioning in plant cells, including monocotyledonous or dicotyledonous plants, or tissues or organs derived from such cells. The promoter may regulate gene expression constitutively, or differentially with respect to the tissue in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, or metal ions, amongst others.

Examples of promoters useful in performing this embodiment include the CaMV 35S promoter, NOS promoter, octopine synthase (OCS) promoter, *Arabidopsis thaliana SSU* gene promoter, napin seed-specific promoter, SCSV promoter, SCBV promoter and the like. In addition to the specific promoters identified herein, cellular promoters for so-called housekeeping genes, including the actin promoters, or promoters of histone-encoding genes, are useful.

The term "terminator" refers to a DNA sequence at the end of a transcriptional unit which signals termination of transcription. Terminators are 3'-non-translated DNA sequences containing a polyadenylation signal, that facilitate the addition of a polyadenylate sequence to the 3'-end of a primary transcript. Terminators active in cells derived from viruses, yeasts, moulds, bacteria, insects, birds, mammals and plants are known and described in the literature. They are isolatable from bacteria, fungi, viruses, animals and/or plants.

Examples of terminators particularly suitable for use in the gene constructs of the present invention include the nopaline synthase (NOS) gene terminator of *Agrobacterium tumefaciens,* the terminator of the Cauliflower mosaic virus (CaMV) 35S gene, the *zein* gene terminator from *Zea mays,* the Rubisco small subunit (SSU) gene terminator sequences and subclover stunt virus (SCSV) gene sequence terminators, amongst others.

Those skilled in the art will be aware of additional promoter sequences and terminator sequences that may be suitable for use in performing the invention. Such sequences may readily be used without any undue experimentation.

Preferably, the gene construct further comprises an origin of replication sequence for its replication in a specific cell type, for example a bacterial cell, when said gene construct is required to be maintained as an episomal genetic element (eg. plasmid or cosmid molecule) in said cell. Preferred origins of replication include, but are not limited to, the *f1*-ori and *col*E1 origins of replication.

Preferably, the gene construct further comprises a selectable marker gene or genes that are functional in a cell into which said gene construct is introduced.

As used herein, the term "selectable marker gene" includes any gene which confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells which are transfected or transformed with a gene construct of the invention or a derivative thereof.

Suitable selectable marker genes contemplated herein include the ampicillin resistance (Amp^r), tetracyclin-resistance gene (Tc^r), bacterial kanamycin resistance gene (Kan^r), phosphinothricin resistance gene, neomycin phosphotransferase gene (*npt*II), hygromycin resistance gene, β-glucuronidase (GUS) gene, chloramphenicol acetyltransferase (CAT) gene, and luciferase gene, amongst others.

In a related embodiment, the invention extends to the use of an isolated gene comprising a nucleotide sequence that is homologous to a protein-encoding region of a gene of *A. thaliana* that is positioned between about 46cM to about 50.74cM on chromosome 2 in the preparation of a gene construct for enhancing the transpiration efficiency of a plant.

In an alternative embodiment of the invention, the transpiration efficiency of a plant is enhanced by classical breeding approaches, comprising introgressing the isolated gene into a plant. For introgression of a gene, the gene is transferred from its native

genetic background into another genetic background using standard breeding, for example, a gene that enhances transpiration efficiency in a progenitor such as a diploid cotton or diploid wheat may be transferred into a commercial tetraploid cotton or hexaploid wheat, respectively, by standard crossing; followed by several generations of back-crossing to remove the genetic background of the progenitor. Naturally, continued selection of the gene of interest is required, such as, for example, facilitated by the use of markers.

A fifth aspect of the present invention provides a plant having enhanced transpiration efficiency, wherein said plant is produced by a method described herein.

The present invention is further described with reference to the following non-limiting examples.

EXAMPLE 1

$^{12}C/^{13}C$ discrimination as a marker for screening genetic variation in transpiration efficiency.

Experimental conditions and sampling procedures were established to allow the control of many factors, other than genetic, that influence transpiration efficiency at the level of individual leaves and plants. These factors fall into several categories: (a) characteristics of the seedling's micro-environment: temperature, light, humidity, boundary layer around the leaves, root growth conditions; (b) developmental and morphological effects that modify gas exchange and C metabolism and therefore carbon isotopic signature (eg age, stage, posture); and (c) seed effects.

We developed high resolution mass-spectrometer techniques for measuring C isotope ratios in whole tissues or carbon compounds such as soluble sugars -ie a measure of integrated transpiration efficiency over the plant's life or over a day, respectively, and also for measuring instantaneous transpiration efficiency during gas exchange.

This means:

- 0.1 per mil analytical precision in the measurement of the isotopic composition of leaf carbon. Discrimination, (–), is approximately the isotope ratio of carbon in source CO_2 minus that of plant organic carbon. In a particular experiment, the source CO_2 is common to all genotypes.
- 0.1 per mil biological precision, that is variation between replicated seedlings, grown in soil, either in growth chambers or in glasshouses with CO_2, humidity and temperature control (corresponding to approximately 1.5% variation in transpiration efficiency).
- The ability to grow and screen large batches of seedlings in glasshouses or growth chambers (up to 1500), under standardised leaf and root growth conditions, to a rosette size of several cm within 2-3 weeks allowing individual measurements, on the same plant, of isotope ratios and also of the underlying properties (eg in situ measurement of leaf temperature by infra-red thermometry as a measure of stomatal conductance; chlorophyll fluorescence; leaf expansion).

EXAMPLE 2

Natural genetic variation in transpiration efficiency in *Arabidopsis thaliana*

A. thaliana ecotypes were screened for leaf Δ under glasshouse conditions. There was a large spread of values (corresponding to approximately 30 % genetic variation in transpiration efficiency). However, large environmental effects were noted. A few contrasted ecotypes were selected at the two extremes of the range of Δ values and compared under various conditions of irradiance (150 to 500 μE $m^{-2}s^{-1}$), light spectrum (Red/Far-Red ratios) and air humidity (60 to 90%) while roots were always well watered. The magnitude of genetic differences in transpiration efficiency was very much influenced by environmental conditions. This was in part due to variations among ecotypes in the dependence of photosynthesis on light and vapour pressure deficit. Genetic differences were maximal under a combination of high light and low humidity, in growth chambers.

The ecotypes *Columbia (Col)* and *Landsberg erecta (*Ld-*er)* have extreme carbon isotope discrimination values, with *Col* always having smaller values than *Ld-er*, and thus a greater transpiration efficiency.

EXAMPLE 3

Identification of a locus associated with transpiration efficiency in *A. thaliana*

Quantitative Trait Loci (QTL) analysis of the Lister and Dean's (1993) Recombinant Inbred Lines (later referred to as RILs) was performed to identify and map a locus associated with carbon isotope discrimination (Δ). The RILs were from a cross between Col-4 and Ler-0. Our analysis confirmed the importance of genes around the *ER* locus, and a role for genes other than *ERECTA* in conferring transpiration efficiency on *A. thaliana*.

More particularly, 300 RI mapping lines between Col and Ler ecotypes, available at the Arabidopsis Stock Centre, were generated from a cross between the *Arabidopsis* ecotypes Columbia (Col4) and Landsberg erecta (Ler-0 carrying *er1*) (Lister and Dean, 1993), using Columbia as the male parent. A subset of 100 of these lines, chosen as the most densely and reliably mapped were used in the present analysis.

The seeds were multiplied in a glasshouse in an attempt to minimize confounding seed effects in our comparisons. Large numbers of seeds were obtained for most lines except for a few, including Col4 parent, which had to be re-ordered following low seed viability of the original sample sent by the Stock Centre. The seeds harvested in these propagation runs were used throughout all our experiments to date.

Loci were analysed using two programs, QTL cartographer and MQTL. These programs compute statistics of a trait at each marker position, using a range of methods [linear regression (LR), stepwise regression (SR), and likelihood approaches (Single interval mapping (SIM) which treats values at individual markers as independent values, and composite interval mapping (CIM) which allows for interactions between markers and associated locus)]. By nature each of these methods has some biases and embedded assumptions, hence the importance of analysing data with more than one program. Only results that were consistent between the two programs, and robust to additions or deletions to the set of background markers used for composite interval mapping are reported below.

Initial QTL analysis was done in parallel to seed multiplication on a subset of 40 lines for which enough seeds were sent. Once all seeds had been multiplied this was

repeated on the full set of 100 lines. These two analyses indicated the existence of a locus for carbon isotope discrimination (Δ), that maps to the region including the *ERECTA* locus on chromosome 2, at approximately 46-51 cM (Table 1, run 1&2).

Given the complexity and integrative nature of Δ as a physiological trait, such a small number of loci associated with the trait was not expected. Subsequent experiments were therefore designed to test these results and assess their stability across the range of environmental conditions known for their effects on gene expression related to Δ (see above). QTL analysis was repeated on several completely independent data sets obtained under highly controlled conditions in glasshouses or growth chambers, where either air humidity, photoperiod or irradiance (amount, diurnal pattern, day to day variation) was varied. Depending on the experiment, all 100 recombinants inbred lines were included or only the subset of lines with cross-overs on chromosome 2. These experiments confirmed that genetic variation in Δ could be mostly ascribed to a portion of chromosome 2 (Table 1) between about 46-50.7 cM.

When RILs were sorted graphically according to carbon isotope discrimination and their genotype at the *ER* marker (50.64 cM) and its vicinity (Ld-*er1* genotype or Col-*ER* genotype), lines which were Ld-*er* at the *ERECTA* marker ranked mostly at the high end of carbon isotope discrimination values. In contrast, lines having a Col-*ERECTA* marker genotype ranked mostly at the low end of carbon isotope discrimination values (data available on request). In the middle of the range of carbon isotope discrimination values, there was some overlap between the two sets of lines. Some lines were always at an extreme (in all 18 experiments performed), while the ranking of other lines was more unstable. These data indicate a locus for transpiration efficiency, as determined by the carbon isotope discrimination value, in the vicinity of the *ERECTA* locus on chromosome 2 (Table 1). This locus most likely involves the *ER* gene. Depending on the positions of cross-overs between *Ld-er* and *Col*, recombination between *ERECTA* and one or more of the other genes influences the transpiration efficiency phenotype of the progeny.

EXAMPLE 5

Determination of a role for the *ERECTA* gene in regulating transpiration efficiency

We compared Col and Ler ecotypes with near-isogenic mutant lines for the *erecta* gene, to examine a possible role of the *ERECTA* gene in determining carbon isotope discrimination (Δ).

Plants expressing the wild type *ERECTA* gene (SEQ ID NO: 1), or an *erecta* mutant allele in the Columbia background (eg. Col-*er101* to -*er105*; or Col-*er108* to -*er123*), have been publicly described. Two of these mutants were available for comparison to the isogenic or near-isogenic lines (Table 2).

Col4, the other parental line (*ER*) for Lister and Dean's RILs was systematically included in the comparison. Where possible, other Col "ecotypes" were also included, (eg. Col0 , Col3-7), to assess their similarity with respect to carbon isotope discrimination, especially compared to the RIL parental ecotype Col4.

The results of these comparisons are described in Table 3. Data indicate that the differences in carbon isotope discrimination values between *er* and *ER* lines for 15 different experimental runs corresponding to growth under low to high light (100 to 800 μE m^{-2} s^{-1}), low to high humidity (40 to 85%), short to long days (8, 10, 24hrs), normal to high temperatures (22/20°C to 28/20 °C).

As expected, the spread of carbon isotope discrimination values among lines varied with environmental conditions. Lines carrying *er* mutations have a greater carbon isotope discrimination value overall than those having the *ER* wild type gene (see Table 3, column 1), indicative of a lower water use-efficiency. There is usually little difference in C isotopic discrimination between the various Col lines, (see the similar averages obtained for columns 2, 3, and 4 in Table 3, wherein *er105* is compared to 3 different Col ecotypes, Col0, Col4 and 3176 or Col1). When present, the *er105* mutant always has the greatest carbon isotope discrimination value of all lines, including *er1* and *er2* (columns 2-4 compared to columns 5-6 in Table 3, or column 8 compared to column 9 in Table 3). The value measured in the *er105* mutant is always significantly greater than in the *ER* isogenic line (column 4 in Table 3). The value measured in *er1* (Landsberg parental line NW20) is usually also greater than that in the *ER* lines 3177 (near isogenic, column 6 of Table 3), and to a lesser extent Col4 (Columbia parental line, column 7 of Table 3). These observations give

direct evidence that the *ERECTA* gene plays a significant role in determining genetic differences in carbon isotopic discrimination in *Arabidopsis*.

This conclusion is independently confirmed by leaf gas exchange measurements that allow the direct measure of transpiration efficiency (ratio of net CO_2 fixation to water loss; column 4 in Table 4; Figures 1a-1c, 2a-2c). Measurements on mature leaves reveal that *ER* lines are characterised by a greater ratio of CO_2 assimilation to water loss than lines carrying *er* mutations. This is most obvious when comparing the pair *Col1/ er105* with a 21% greater transpiration efficiency (ratio A/E) in *Col1* than *er105*, or the pair *Col1/er2* with a 16% greater transpiration efficiency in *Col1*. Consistent with the measurements of carbon isotope discrimination, the effect *er/ER* is relatively smaller in the *Ld* background (9% greater ratio A/E in *Ld-ER* (3177) than the *Ld-er1*(NSW20) background.

Also consistent with the carbon discrimination measurements, is the 20% difference in transpiration efficiency between the two RILs parental lines (4.06 and 3.38 mmolC/molH2O in *Col4-ER* and *Ld-er1*, respectively).

The fact that of all 3 *erecta* mutants examined, *er105* has the most extreme carbon discrimination and transpiration efficiency phenotypes suggests that the *er105* mutation affects a more crucial part of the *ERECTA* gene than *er2* or *er1*. This is consistent with the published data on the *er105* mutant. This mutation corresponds to the insertion of a large "foreign insert" in the *ERECTA* gene (1200bp). The insertion totally inhibits transcription of the gene and causes the strongest *erecta* phenotype of all *erecta* mutants isolated in Col (with respect to inflorescence clustering and silique width and shape. Alternatively, or in addition, data indicate that *erecta* mutations have a stronger effect on carbon isotope discrimination values in a Columbia genetic background than in a Landsberg background (comparison of phenotypic effects of *er105* and *er1*), implying that other genes, polymorphic between Landsberg and Columbia ecotypes, interact with *ERECTA* in determining transpiration efficiency. This could also account for the greater difference in transpiration efficiency between *er/ER* lines in Col background than in a Ld background (see above, Table 4). Alternatively, or in addition, data indicate that the *erecta* mutation is not the only mutation present in the *er105* mutant. For example,

the mutagenized Col seeds may have carried the *gl1* mutation, induced by the fast neutron irradiation, that also contributes to the phenotype observed.

A comparison of transcript profiles in *er/ER* isogenic lines (in both Col and Ld background) allows determination of the involvement of additional genes to *ERECTA* and the effect of environment on their expression.

TABLE 1

QTL Analysis of Carbon Isotope Discrimination in Lister and Dean's Recombinant Inbred Lines

RUN No.	chr2 locus	QTL	chr4 locus	CONCLUSION	
Experimental conditions		**analysis method**		**QTLs number**	**predicted map position**
Run 1 (40 lines) Glasshouse-					
12h day length	**58.5**	SIM&CIM		2	chr2: 58.5-61.02
irradiance150-350 μE m^{-2} s^{-1}	46.77	SIM&CIM			chr2: 46.77-50.75
Seedlings transferred from	**61.02**	SIM	108.5		
agar plates					
Run1 data					
but with using different					
markers					
	56.94 to 58.00	CIM&SIM		**1**	
	46.77 to 50.75	SIM			
	63.02				

45

RUN No.	chr2 locus	QTL	chr4 locus	CONCLUSION	
Experimental conditions		**analysis method**		**QTLs number**	**predicted map position**
Run 1					
with different number of lines					
	58.5 to 61.02			2	
	56-61				
Run2					
Glasshouse					
September					
from seeds sown on soil					
batch 1	50.75	CIM (QTL cart)		2	chr2: 56.94-61.02
	61.02	MQTL			chr2: 50.75
batch 2	?50.75	MQTL		NS	
batch3-5					
all batches	58.5	MQTLcart		NS	
	56.94-58.5	MQTL			

RUN No.	chr2 locus	QTL	chr4 locus	CONCLUSION	
Experimental conditions		analysis method		QTLs number	predicted map position
Run 3					
37 lines: parents and lines with crossing-overs on chromosome 2					
5 growth conditions differing in humidity, irradiance, mode of establishment (seeds sown on soil or seedlings transplanted from agar)					
batch B	61.02-61.06		108 NS		
batch C	56.94-58.00				
batch D	63.02	QTLcar			
	63.02	MQTL			
all batches (conditions)	58.5			1or 2?	chr2: 56.94-58.5
	61.02			3	chr2: 61.02-63.02

47

RUN No.	chr2 locus	QTL	chr4 locus	CONCLUSION	
Experimental conditions		analysis method		QTLs number	predicted map position
Run 4					
same lines as Run 3					
growth chambers					
10h daylight					
	50.74				chr2: 50.74
Run 5					
repeat of run 1 BUT ALL					
lines					
	50.74			1	chr2: 50.74
Run 7					
same lines as in run 1 but					
in growth chamber					
and higher light					
10h daylength	46.77-50.75	CIM&SIM		1	chr 2: 46.77-5065
470-510 μE m^{-2} s^{-1} irradiance					

TABLE 2

Background	Mutation	Stock Centre name	Isogenic *ER* line and Stock Centre Name
Landsberg	*er1*	CS20 or NW20 [a]	3177 or CS163
Columbia	*er2*[b]	3401	Col1 or 3176
Columbia	*er105* [c]		Col3 with gl1 marker or Col0

a, NW20 is an *Ler* parent for Lister and Dean's recombinant lines, carrying the *er1* mutation. Lines 3177 or CS163 are the closest isogenic ER lines.

b, *er2* is an *er* allele identified by Rédéi in Col background. Col1 or 3176 are the closest Col near-isogenic lines.

c, *er105* was isolated from a fast-neutron-irradiated Col seed population (Torii *et al.*, 1996).

d, Col4, the *Col* parent for the Lister and Dean's parent was systamically included in all comparisons.

TABLE 3

Comparison of er/ER lines in both Col and Ld background for carbon isotope discrimination values (per ml) in leaf material under a range of environmental conditions

Run No.	Differences in mean carbon isotope discrimination values (per mil)								
	(1) *er-ER (all lines)*	(2) er105-Col0	(3) er105-Col4	(4) er105-3176	(5) er2-3176	(6) er1-3177	(7) er1-Col4 (parental lines for RILs)	(8) er105-Col*i*	(9) er1-Col*i*
1	0.13	0.16						0.16	
2	0.89	1.18						1.18	
3	0.26	0.11						0.11	
4	1.12	1.60						1.60	
5	1.03		1.83	1.67	0.92	0.64	0.82	1.75	0.73
6	0.70	1.13	1.01	0.71	0.27	0.74	0.73	0.95	0.73
7	0.70	1.32	1.12	1.23	0.75	0.35	0.05	1.22	0.15
8	0.59	1.16	1.11	1.19	0.54	0.28	0.06	1.16	0.17
9	0.30	1.09	0.77	0.77	0.02	0.00	0.56	0.88	0.28
10	0.56	1.05	0.94	0.87	0.38	0.39	0.33	0.95	0.36

11	0.48		0.40				0.52	0.40	0.52
12	0.36	0.82	1.31	1.08	0.33	0.05	0.07	1.07	0.01
13	0.38		0.90	0.82	0.07	0.60	0.52	0.86	0.56
14	0.65	1.42	0.60			0.58	0.06	1.01	0.32
15	0.82						0.82		0.82
For all runs:									
Mean	0.60	1.01	1.00	1.04	0.41	0.40	0.41	0.95	0.42
S.E.	0.07	0.14	0.11	0.11	0.10	0.08	0.09	0.12	0.08
For Common runs:									
Mean:	0.58	1.10	1.12	1.04	0.41	0.38	0.39	1.11	0.37
S.E.	0.08	0.05	0.11	0.11	0.10	0.09	0.10	0.10	0.09

TABLE 4

Run 9- December 2001: Leaf gas exchange measurements in er/ER Arabidopsis lines

	Genotype		(1) E mmol H_2O /m^2/s	(2) A	(3) Gw mol/m^2/s	(4) A/E mmolC/	(5) pa µbar	(6) pi µbar	(7) pi/pa	(8) 1-pi/pa
				µmolC/m^2/s		molH$_2$O				
Row (1) **Ld-ER**	**3177-G**	Mean	3.38	12.33	0.273	3.67	360	282	0.782	0.218
		S.E.	0.48	1.64	0.039	0.14	10	11	0.010	0.010
Row (2) **Ld-er**	**NSW20 E**	Mean	2.59	8.73	0.218	3.38	348	280	0.804	0.196
		S.E.	0.07	0.31	0.005	0.04	5	4	0.002	0.002
Row (3) **Col-ER**	**933**	Mean	3.41	13.55	0.291	4.06	350	270	0.772	0.228
		S.E.	0.40	1.16	0.040	0.22	4	7	0.020	0.020
Row (4) **Col-ER**	**3176 J**	Mean	2.23	10.13	0.180	4.55	346	254	0.734	0.266
	(Col1)	S.E.	0.50	1.47	0.048	0.24	5	9	0.021	0.021

Row (5) Col-er	er105 F	Mean	2.27	8.55	0.198	3.76	356	283	0.795	0.205
		S.E.	0.03	0.17	0.005	0.07	11	10	0.006	0.006
Row (6) Col-er	er2C	Mean	3.06	11.90	0.256	3.92	357	279	0.780	0.220
		S.E.	0.22	0.56	0.027	0.12	1	6	0.014	0.014

CONCLUSION

Comparison Ld-ER/Ld-er	er line has lower A/E with lower g and lower A
	The difference in A/E is driven by A
Comparison 933/NSW20	NSW20 (er) has lower A/E with lower g and lower A
Comparison Col1/Ld-er1	The difference in A/E is driven by A
Comparison Col1/Col-er105	er105 has MUCH lower A/E with Higher g and lower A
	i.e. the difference in A/E is driven by A and g
Comparison Col1/Col-er2	er2 has lower A/E with MUCH higher g and HIGHER A
	i.e. the difference in A/E is driven by g and is opposed or not driven by A

NOTE: p_a and p_i are the ambient and intercellular partial pressures of CO_2, respectively.

SEQUENCE LISTING

<110> The Australian National University

Masle, Josette

Farquhar, Graham

Gilmore, Scott

<120> Method of producing plants having enhanced transpiration efficiency and plants produced therefrom I

<130> 500592/MRO

<160> 10

<170> PatentIn version 3.1

<210> 1

<211> 3176

<212> DNA

<213> Arabidopsis thaliana ERECTA allele

<220>

<221> CDS

<222> (77)..(3004)

<223>

<400> 1

gtttcttctt catggagact tgaaagcttt taaagtatat ctaaaaacgc agtcgtttta 60

```
agactgtgtg tgagaa atg gct ctg ttt aga gat att gtt ctt ctt ggg ttt
112
                  Met Ala Leu Phe Arg Asp Ile Val Leu Leu Gly Phe
                  1               5                  10
5
ctc ttc tgc ttg agc tta gta gct act gtg act tca gag gag gga gca
160
Leu Phe Cys Leu Ser Leu Val Ala Thr Val Thr Ser Glu Glu Gly Ala
        15                  20                  25
10
acg ttg ctg gag att aag aag tca ttc aaa gat gtg aac aat gtt ctt
208
Thr Leu Leu Glu Ile Lys Lys Ser Phe Lys Asp Val Asn Asn Val Leu
    30                  35                  40
15
tat gac tgg aca act tca cct tct tcg gat tat tgt gtc tgg aga ggt
256
Tyr Asp Trp Thr Thr Ser Pro Ser Ser Asp Tyr Cys Val Trp Arg Gly
45                  50                  55                  60
20
gtg tct tgt gaa aat gtc acc ttc aat gtt gtt gct ctt aat ttg tca
304
Val Ser Cys Glu Asn Val Thr Phe Asn Val Val Ala Leu Asn Leu Ser
                65                  70                  75
25
gat ttg aat ctt gat gga gaa atc tca cct gct att gga gat ctc aag
352
Asp Leu Asn Leu Asp Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Lys
            80                  85                  90
30
agt ctc ttg tca att gat ctg cga ggt aat cgc ttg tct gga caa atc
400
Ser Leu Leu Ser Ile Asp Leu Arg Gly Asn Arg Leu Ser Gly Gln Ile
        95                  100                 105
35
cct gat gag att ggt gac tgt tct tct ttg caa aac tta gac tta tcc
448
Pro Asp Glu Ile Gly Asp Cys Ser Ser Leu Gln Asn Leu Asp Leu Ser
    110                 115                 120
40
ttc aat gaa tta agt ggt gac ata ccg ttt tcg att tcg aag ttg aag
496
Phe Asn Glu Leu Ser Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys
125                 130                 135                 140
45
caa ctt gag cag ctg att ctg aag aat aac caa ttg ata gga ccg atc
544
Gln Leu Glu Gln Leu Ile Leu Lys Asn Asn Gln Leu Ile Gly Pro Ile
                145                 150                 155
50
cct tca aca ctt tca cag att cca aac ctg aaa att ctg gac ttg gca
592
Pro Ser Thr Leu Ser Gln Ile Pro Asn Leu Lys Ile Leu Asp Leu Ala
            160                 165                 170
55
cag aat aaa ctc agt ggt gag ata cca aga ctt att tac tgg aat gaa
640
```

```
Gln Asn Lys Leu Ser Gly Glu Ile Pro Arg Leu Ile Tyr Trp Asn Glu
        175                 180                 185

gtt ctt cag tat ctt ggg ttg cga gga aac aac tta gtc ggt aac att
688
Val Leu Gln Tyr Leu Gly Leu Arg Gly Asn Asn Leu Val Gly Asn Ile
    190                 195                 200

tct cca gat ttg tgt caa ctg act ggt ctt tgg tat ttt gac gta aga
736
Ser Pro Asp Leu Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg
205                 210                 215                 220

aac aac agt ttg act ggt agt ata cct gag acg ata gga aat tgc act
784
Asn Asn Ser Leu Thr Gly Ser Ile Pro Glu Thr Ile Gly Asn Cys Thr
                225                 230                 235

gcc ttc cag gtt ttg gac ttg tcc tac aat cag cta act ggt gag atc
832
Ala Phe Gln Val Leu Asp Leu Ser Tyr Asn Gln Leu Thr Gly Glu Ile
            240                 245                 250

cct ttt gac atc ggc ttc ctg caa gtt gca aca tta tca ttg caa ggc
880
Pro Phe Asp Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly
        255                 260                 265

aat caa ctc tct ggg aag att cca tca gtg att ggt ctc atg caa gcc
928
Asn Gln Leu Ser Gly Lys Ile Pro Ser Val Ile Gly Leu Met Gln Ala
    270                 275                 280

ctt gca gtc tta gat cta agt ggc aac ttg ttg agt gga tct att cct
976
Leu Ala Val Leu Asp Leu Ser Gly Asn Leu Leu Ser Gly Ser Ile Pro
285                 290                 295                 300

ccg att ctc gga aat ctt act ttc acc gag aaa ttg tat ttg cac agt
1024
Pro Ile Leu Gly Asn Leu Thr Phe Thr Glu Lys Leu Tyr Leu His Ser
                305                 310                 315

aac aag ctg act ggt tca att cca cct gag ctt gga aac atg tca aaa
1072
Asn Lys Leu Thr Gly Ser Ile Pro Pro Glu Leu Gly Asn Met Ser Lys
            320                 325                 330

ctc cat tac ctg gaa ctc aat gat aat cat ctc acg ggt cat ata cca
1120
Leu His Tyr Leu Glu Leu Asn Asp Asn His Leu Thr Gly His Ile Pro
        335                 340                 345

cca gag ctt ggg aag ctt act gac ttg ttt gat ctg aat gtg gcc aac
1168
Pro Glu Leu Gly Lys Leu Thr Asp Leu Phe Asp Leu Asn Val Ala Asn
    350                 355                 360
```

```
aat gat ctg gaa gga cct ata cct gat cat ctg agc tct tgc aca aat
1216
Asn Asp Leu Glu Gly Pro Ile Pro Asp His Leu Ser Ser Cys Thr Asn
365                 370                 375                 380

cta aac agc tta aat gtt cat ggg aac aag ttt agt ggc act ata ccc
1264
Leu Asn Ser Leu Asn Val His Gly Asn Lys Phe Ser Gly Thr Ile Pro
                385                 390                 395

cga gca ttt caa aag cta gaa agt atg act tac ctt aat ctg tcc agc
1312
Arg Ala Phe Gln Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu Ser Ser
            400                 405                 410

aac aat atc aaa ggt cca atc ccg gtt gag cta tct cgt atc ggt aac
1360
Asn Asn Ile Lys Gly Pro Ile Pro Val Glu Leu Ser Arg Ile Gly Asn
        415                 420                 425

tta gat aca ttg gat ctt tcc aac aac aag ata aat gga atc att cct
1408
Leu Asp Thr Leu Asp Leu Ser Asn Asn Lys Ile Asn Gly Ile Ile Pro
    430                 435                 440

tct tcc ctt ggt gat ttg gag cat ctt ctc aag atg aac ttg agt aga
1456
Ser Ser Leu Gly Asp Leu Glu His Leu Leu Lys Met Asn Leu Ser Arg
445                 450                 455                 460

aat cat ata act ggt gta gtt cca ggc gac ttt gga aat cta aga agc
1504
Asn His Ile Thr Gly Val Val Pro Gly Asp Phe Gly Asn Leu Arg Ser
                465                 470                 475

atc atg gaa ata gat ctt tca aat aat gat atc tct ggc cca att cca
1552
Ile Met Glu Ile Asp Leu Ser Asn Asn Asp Ile Ser Gly Pro Ile Pro
            480                 485                 490

gaa gag ctt aac caa tta cag aac ata att ttg ctg aga ctg gaa aat
1600
Glu Glu Leu Asn Gln Leu Gln Asn Ile Ile Leu Leu Arg Leu Glu Asn
        495                 500                 505

aat aac ctg act ggt aat gtt ggt tca tta gcc aac tgt ctc agt ctc
1648
Asn Asn Leu Thr Gly Asn Val Gly Ser Leu Ala Asn Cys Leu Ser Leu
    510                 515                 520

act gta ttg aat gta tct cat aac aac ctc gta ggt gat atc cct aag
1696
Thr Val Leu Asn Val Ser His Asn Asn Leu Val Gly Asp Ile Pro Lys
525                 530                 535                 540
```

```
aac aat aac ttc tca aga ttt tca cca gac agc ttc att ggc aat cct
1744
Asn Asn Asn Phe Ser Arg Phe Ser Pro Asp Ser Phe Ile Gly Asn Pro
                545                 550                 555

ggt ctt tgc ggt agt tgg cta aac tca ccg tgt cat gat tct cgt cga
1792
Gly Leu Cys Gly Ser Trp Leu Asn Ser Pro Cys His Asp Ser Arg Arg
            560                 565                 570

act gta cga gtg tca atc tct aga gca gct att ctt gga ata gct att
1840
Thr Val Arg Val Ser Ile Ser Arg Ala Ala Ile Leu Gly Ile Ala Ile
        575                 580                 585

ggg gga ctt gtg atc ctt ctc atg gtc tta ata gca gct tgc cga ccg
1888
Gly Gly Leu Val Ile Leu Leu Met Val Leu Ile Ala Ala Cys Arg Pro
    590                 595                 600

cat aat cct cct cct ttt ctt gat gga tca ctt gac aaa cca gta act
1936
His Asn Pro Pro Pro Phe Leu Asp Gly Ser Leu Asp Lys Pro Val Thr
605                 610                 615                 620

tat tcg aca ccg aag ctc gtc atc ctt cat atg aac atg gca ctc cac
1984
Tyr Ser Thr Pro Lys Leu Val Ile Leu His Met Asn Met Ala Leu His
                625                 630                 635

gtt tac gag gat atc atg aga atg aca gag aat cta agt gag aag tat
2032
Val Tyr Glu Asp Ile Met Arg Met Thr Glu Asn Leu Ser Glu Lys Tyr
            640                 645                 650

atc att ggg cac gga gca tca agc act gta tac aaa tgt gtt ttg aag
2080
Ile Ile Gly His Gly Ala Ser Ser Thr Val Tyr Lys Cys Val Leu Lys
        655                 660                 665

aat tgt aaa ccg gtt gcg att aag cgg ctt tac tct cac aac cca cag
2128
Asn Cys Lys Pro Val Ala Ile Lys Arg Leu Tyr Ser His Asn Pro Gln
    670                 675                 680

tca atg aaa cag ttt gaa aca gaa ctc gag atg cta agt agc atc aag
2176
Ser Met Lys Gln Phe Glu Thr Glu Leu Glu Met Leu Ser Ser Ile Lys
685                 690                 695                 700

cac aga aat ctt gtg agc cta caa gct tat tcc ctc tct cac ttg ggg
2224
His Arg Asn Leu Val Ser Leu Gln Ala Tyr Ser Leu Ser His Leu Gly
                705                 710                 715

agt ctt ctg ttc tat gac tat ttg gaa aat ggt agc ctc tgg gat ctt
2272
```

```
Ser Leu Leu Phe Tyr Asp Tyr Leu Glu Asn Gly Ser Leu Trp Asp Leu
            720                 725                 730

ctt cat ggc cct acg aag aaa aag act ctt gat tgg gac aca cgg ctt
2320
Leu His Gly Pro Thr Lys Lys Lys Thr Leu Asp Trp Asp Thr Arg Leu
        735                 740                 745

aag ata gca tat ggt gca gca caa ggt tta gct tat cta cac cat gac
2368
Lys Ile Ala Tyr Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp
    750                 755                 760

tgt agt cca agg atc att cac aga gac gtg aag tcg tcc aac att ctc
2416
Cys Ser Pro Arg Ile Ile His Arg Asp Val Lys Ser Ser Asn Ile Leu
765                 770                 775                 780

ttg gac aaa gac tta gag gct cgt ttg aca gat ttt gga ata gcg aaa
2464
Leu Asp Lys Asp Leu Glu Ala Arg Leu Thr Asp Phe Gly Ile Ala Lys
                785                 790                 795

agc ttg tgt gtg tca aag tca cat act tca act tac gtg atg ggc acg
2512
Ser Leu Cys Val Ser Lys Ser His Thr Ser Thr Tyr Val Met Gly Thr
            800                 805                 810

ata ggt tac ata gac ccc gag tat gct cgc act tca cgg ctc act gag
2560
Ile Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Thr Glu
        815                 820                 825

aaa tcc gat gtc tac agt tat gga ata gtc ctt ctt gag ttg tta acc
2608
Lys Ser Asp Val Tyr Ser Tyr Gly Ile Val Leu Leu Glu Leu Leu Thr
    830                 835                 840

cga agg aaa gcc gtt gat gac gaa tcc aat ctc cac cat ctg ata atg
2656
Arg Arg Lys Ala Val Asp Asp Glu Ser Asn Leu His His Leu Ile Met
845                 850                 855                 860

tca aag acg ggg aac aat gaa gtg atg gaa atg gca gat cca gac atc
2704
Ser Lys Thr Gly Asn Asn Glu Val Met Glu Met Ala Asp Pro Asp Ile
                865                 870                 875

aca tcg acg tgt aaa gat ctc ggt gtg gtg aag aaa gtt ttc caa ctg
2752
Thr Ser Thr Cys Lys Asp Leu Gly Val Val Lys Lys Val Phe Gln Leu
            880                 885                 890

gca ctc cta tgc acc aaa aga cag ccg aat gat cga ccc aca atg cac
2800
Ala Leu Leu Cys Thr Lys Arg Gln Pro Asn Asp Arg Pro Thr Met His
        895                 900                 905
```

cag gtg act cgt gtt ctc ggc agt ttt atg cta tcg gaa caa cca cct
2848
Gln Val Thr Arg Val Leu Gly Ser Phe Met Leu Ser Glu Gln Pro Pro
910 915 920

gct gcg act gac acg tca gcg acg ctg gct ggt tcg tgc tac gtc gat
2896
Ala Ala Thr Asp Thr Ser Ala Thr Leu Ala Gly Ser Cys Tyr Val Asp
925 930 935 940

gag tat gca aat ctc aag act cct cat tct gtc aat tgc tct tcc atg
2944
Glu Tyr Ala Asn Leu Lys Thr Pro His Ser Val Asn Cys Ser Ser Met
945 950 955

agt gct tct gat gct caa ctg ttt ctt cgg ttt gga caa gtt att tct
2992
Ser Ala Ser Asp Ala Gln Leu Phe Leu Arg Phe Gly Gln Val Ile Ser
960 965 970

cag aac agt gag tagtttttcg ttaggaggag aatctttaaa acggtatctt
3044
Gln Asn Ser Glu
975

ttcgttgcgt taagctgtta gaaaaattaa tgtctcatgt aaagtattat gcactgcctt
3104

attattatta gacaagtgtg tggtgtgaat atgtcttcag actggcactt agacttccta
3164

taagttcttg cc
3176

<210> 2

<211> 976

<212> PRT

<213> Arabidopsis thaliana ERECTA allele

<400> 2

Met Ala Leu Phe Arg Asp Ile Val Leu Leu Gly Phe Leu Phe Cys Leu
1 5 10 15

Ser Leu Val Ala Thr Val Thr Ser Glu Glu Gly Ala Thr Leu Leu Glu
20 25 30

Ile Lys Lys Ser Phe Lys Asp Val Asn Asn Val Leu Tyr Asp Trp Thr

35 40 45

Thr Ser Pro Ser Ser Asp Tyr Cys Val Trp Arg Gly Val Ser Cys Glu
50 55 60

Asn Val Thr Phe Asn Val Val Ala Leu Asn Leu Ser Asp Leu Asn Leu
65 70 75 80

Asp Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Lys Ser Leu Leu Ser
85 90 95

Ile Asp Leu Arg Gly Asn Arg Leu Ser Gly Gln Ile Pro Asp Glu Ile
100 105 110

Gly Asp Cys Ser Ser Leu Gln Asn Leu Asp Leu Ser Phe Asn Glu Leu
115 120 125

Ser Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln Leu Glu Gln
130 135 140

Leu Ile Leu Lys Asn Asn Gln Leu Ile Gly Pro Ile Pro Ser Thr Leu
145 150 155 160

Ser Gln Ile Pro Asn Leu Lys Ile Leu Asp Leu Ala Gln Asn Lys Leu
165 170 175

Ser Gly Glu Ile Pro Arg Leu Ile Tyr Trp Asn Glu Val Leu Gln Tyr
180 185 190

Leu Gly Leu Arg Gly Asn Asn Leu Val Gly Asn Ile Ser Pro Asp Leu
195 200 205

Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Asn Asn Ser Leu
210 215 220

Thr Gly Ser Ile Pro Glu Thr Ile Gly Asn Cys Thr Ala Phe Gln Val
225 230 235 240

Leu Asp Leu Ser Tyr Asn Gln Leu Thr Gly Glu Ile Pro Phe Asp Ile
245 250 255

Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Gln Leu Ser
260 265 270

Gly Lys Ile Pro Ser Val Ile Gly Leu Met Gln Ala Leu Ala Val Leu
275 280 285

Asp Leu Ser Gly Asn Leu Leu Ser Gly Ser Ile Pro Pro Ile Leu Gly
290 295 300

Asn Leu Thr Phe Thr Glu Lys Leu Tyr Leu His Ser Asn Lys Leu Thr
305 310 315 320

Gly Ser Ile Pro Pro Glu Leu Gly Asn Met Ser Lys Leu His Tyr Leu
325 330 335

Glu Leu Asn Asp Asn His Leu Thr Gly His Ile Pro Pro Glu Leu Gly
340 345 350

Lys Leu Thr Asp Leu Phe Asp Leu Asn Val Ala Asn Asn Asp Leu Glu
355 360 365

Gly Pro Ile Pro Asp His Leu Ser Ser Cys Thr Asn Leu Asn Ser Leu
370 375 380

Asn Val His Gly Asn Lys Phe Ser Gly Thr Ile Pro Arg Ala Phe Gln
385 390 395 400

Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu Ser Ser Asn Asn Ile Lys
405 410 415

Gly Pro Ile Pro Val Glu Leu Ser Arg Ile Gly Asn Leu Asp Thr Leu
420 425 430

Asp Leu Ser Asn Asn Lys Ile Asn Gly Ile Ile Pro Ser Ser Leu Gly
435 440 445

Asp Leu Glu His Leu Leu Lys Met Asn Leu Ser Arg Asn His Ile Thr
450 455 460

Gly Val Val Pro Gly Asp Phe Gly Asn Leu Arg Ser Ile Met Glu Ile
465 470 475 480

Asp Leu Ser Asn Asn Asp Ile Ser Gly Pro Ile Pro Glu Glu Leu Asn
485 490 495

```
Gln Leu Gln Asn Ile Ile Leu Leu Arg Leu Glu Asn Asn Asn Leu Thr
               500                 505                 510

Gly Asn Val Gly Ser Leu Ala Asn Cys Leu Ser Leu Thr Val Leu Asn
          515                 520                 525

Val Ser His Asn Asn Leu Val Gly Asp Ile Pro Lys Asn Asn Asn Phe
      530                 535                 540

Ser Arg Phe Ser Pro Asp Ser Phe Ile Gly Asn Pro Gly Leu Cys Gly
545                 550                 555                 560

Ser Trp Leu Asn Ser Pro Cys His Asp Ser Arg Arg Thr Val Arg Val
                    565                 570                 575

Ser Ile Ser Arg Ala Ala Ile Leu Gly Ile Ala Ile Gly Gly Leu Val
               580                 585                 590

Ile Leu Leu Met Val Leu Ile Ala Ala Cys Arg Pro His Asn Pro Pro
          595                 600                 605

Pro Phe Leu Asp Gly Ser Leu Asp Lys Pro Val Thr Tyr Ser Thr Pro
      610                 615                 620

Lys Leu Val Ile Leu His Met Asn Met Ala Leu His Val Tyr Glu Asp
625                 630                 635                 640

Ile Met Arg Met Thr Glu Asn Leu Ser Glu Lys Tyr Ile Ile Gly His
                    645                 650                 655

Gly Ala Ser Ser Thr Val Tyr Lys Cys Val Leu Lys Asn Cys Lys Pro
               660                 665                 670

Val Ala Ile Lys Arg Leu Tyr Ser His Asn Pro Gln Ser Met Lys Gln
          675                 680                 685

Phe Glu Thr Glu Leu Glu Met Leu Ser Ser Ile Lys His Arg Asn Leu
      690                 695                 700

Val Ser Leu Gln Ala Tyr Ser Leu Ser His Leu Gly Ser Leu Leu Phe
705                 710                 715                 720
```

```
Tyr Asp Tyr Leu Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro
                725                 730                 735

Thr Lys Lys Lys Thr Leu Asp Trp Asp Thr Arg Leu Lys Ile Ala Tyr
            740                 745                 750

Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Ser Pro Arg
        755                 760                 765

Ile Ile His Arg Asp Val Lys Ser Ser Asn Ile Leu Leu Asp Lys Asp
    770                 775                 780

Leu Glu Ala Arg Leu Thr Asp Phe Gly Ile Ala Lys Ser Leu Cys Val
785                 790                 795                 800

Ser Lys Ser His Thr Ser Thr Tyr Val Met Gly Thr Ile Gly Tyr Ile
                805                 810                 815

Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Thr Glu Lys Ser Asp Val
            820                 825                 830

Tyr Ser Tyr Gly Ile Val Leu Leu Glu Leu Leu Thr Arg Arg Lys Ala
        835                 840                 845

Val Asp Asp Glu Ser Asn Leu His His Leu Ile Met Ser Lys Thr Gly
    850                 855                 860

Asn Asn Glu Val Met Glu Met Ala Asp Pro Asp Ile Thr Ser Thr Cys
865                 870                 875                 880

Lys Asp Leu Gly Val Val Lys Lys Val Phe Gln Leu Ala Leu Leu Cys
                885                 890                 895

Thr Lys Arg Gln Pro Asn Asp Arg Pro Thr Met His Gln Val Thr Arg
            900                 905                 910

Val Leu Gly Ser Phe Met Leu Ser Glu Gln Pro Pro Ala Ala Thr Asp
        915                 920                 925

Thr Ser Ala Thr Leu Ala Gly Ser Cys Tyr Val Asp Glu Tyr Ala Asn
    930                 935                 940

Leu Lys Thr Pro His Ser Val Asn Cys Ser Ser Met Ser Ala Ser Asp
```

945 950 955 960

Ala Gln Leu Phe Leu Arg Phe Gly Gln Val Ile Ser Gln Asn Ser Glu
965 970 975

<210> 3

<211> 3000

<212> DNA

<213> rice ERECTA homolog

<220>

<221> CDS

<222> (1)..(2997)

<223>

<400> 3

atg gcg gcg gcg agg gcg ccg tgg ctg tgg tgg tgg gtg gtg gtg gtt 48
Met Ala Ala Ala Arg Ala Pro Trp Leu Trp Trp Trp Val Val Val Val
1 5 10 15

gtt ggt gtg gcg gtg gcg gag gcg gcc tcc gga gga gga gga ggg gga 96
Val Gly Val Ala Val Ala Glu Ala Ala Ser Gly Gly Gly Gly Gly Gly
20 25 30

gat ggg gag ggg aag gcg ctg atg ggc gtg aag gcc ggt ttc ggg aac 144
Asp Gly Glu Gly Lys Ala Leu Met Gly Val Lys Ala Gly Phe Gly Asn
35 40 45

gcg gcc aac gcg ctc gtc gac tgg gac ggc ggc gcc gac cac tgc gcg 192
Ala Ala Asn Ala Leu Val Asp Trp Asp Gly Gly Ala Asp His Cys Ala
50 55 60

tgg cgc ggc gtc acc tgc gac aac gcc tcc ttc gcc gtc ctc gcc ctg 240
Trp Arg Gly Val Thr Cys Asp Asn Ala Ser Phe Ala Val Leu Ala Leu
65 70 75 80

aac ttg tca aat cta aac cta gga ggt gag atc tcg ccg gcc atc gga 288
Asn Leu Ser Asn Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly
85 90 95

```
gag ctc aag aat cta cag ttc gtt gat ctc aag ggg aac aag ctc act
336
Glu Leu Lys Asn Leu Gln Phe Val Asp Leu Lys Gly Asn Lys Leu Thr
               100                 105                 110

ggc caa atc cca gat gag att ggg gac tgc atc tcc tta aaa tat ttg
384
Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys Ile Ser Leu Lys Tyr Leu
          115                 120                 125

gat ttg tct ggc aac ttg ctg tat gga gac atc ccc ttc tcc atc tcc
432
Asp Leu Ser Gly Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser
     130                 135                 140

aag ctc aag cag ctt gag gag ctg att ttg aag aac aac cag ctc acg
480
Lys Leu Lys Gln Leu Glu Glu Leu Ile Leu Lys Asn Asn Gln Leu Thr
145                 150                 155                 160

gga ccc atc cct tcc aca ttg tcc caa att cca aat ctc aag aca ttg
528
Gly Pro Ile Pro Ser Thr Leu Ser Gln Ile Pro Asn Leu Lys Thr Leu
                165                 170                 175

gac ctg gca cag aac cag ctt aca ggc gat atc cca agg ctc ata tac
576
Asp Leu Ala Gln Asn Gln Leu Thr Gly Asp Ile Pro Arg Leu Ile Tyr
               180                 185                 190

tgg aat gaa gtt ctg caa tac cta ggt ttg agg ggt aac tca ctg act
624
Trp Asn Glu Val Leu Gln Tyr Leu Gly Leu Arg Gly Asn Ser Leu Thr
          195                 200                 205

gga act ttg tca cct gac atg tgc caa ctg act ggc ctg tgg tac ttt
672
Gly Thr Leu Ser Pro Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe
     210                 215                 220

gat gta agg gga aac aat ctc aca ggg acc att cca gag agc ata ggg
720
Asp Val Arg Gly Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly
225                 230                 235                 240

aac tgc acc agc ttt gag att ctg gac att tcg tat aac caa atc tct
768
Asn Cys Thr Ser Phe Glu Ile Leu Asp Ile Ser Tyr Asn Gln Ile Ser
                245                 250                 255

gga gaa ata cct tac aac ata ggc ttt ctt caa gta gcc aca ctg tca
816
Gly Glu Ile Pro Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser
               260                 265                 270
```

```
ctt caa gga aat aga ctg act ggg aaa att cca gat gtg att ggc ctg
864
Leu Gln Gly Asn Arg Leu Thr Gly Lys Ile Pro Asp Val Ile Gly Leu
        275                 280                 285

atg caa gct ctt gct gtt cta gac ctg agt gag aac gag ctg gta ggg
912
Met Gln Ala Leu Ala Val Leu Asp Leu Ser Glu Asn Glu Leu Val Gly
    290                 295                 300

ccc att cct tct ata ctg ggc aat cta tcc tat act gga aaa cta tat
960
Pro Ile Pro Ser Ile Leu Gly Asn Leu Ser Tyr Thr Gly Lys Leu Tyr
305                 310                 315                 320

tta cat ggg aac aaa ctt act gga gtc ata ccg ccg gag ctt ggg aac
1008
Leu His Gly Asn Lys Leu Thr Gly Val Ile Pro Pro Glu Leu Gly Asn
                325                 330                 335

atg agt aaa ctt agc tac cta caa ctg aat gat aat gaa ttg gtg ggc
1056
Met Ser Lys Leu Ser Tyr Leu Gln Leu Asn Asp Asn Glu Leu Val Gly
            340                 345                 350

aca att cca gca gag ctt ggc aaa ctt gaa gag ctt ttt gaa cta aat
1104
Thr Ile Pro Ala Glu Leu Gly Lys Leu Glu Glu Leu Phe Glu Leu Asn
        355                 360                 365

ctt gcc aac aac aat ctt caa ggt cct att cct gca aac atc agt tct
1152
Leu Ala Asn Asn Asn Leu Gln Gly Pro Ile Pro Ala Asn Ile Ser Ser
    370                 375                 380

tgc act gct cta aac aaa ttc aat gtt tat ggc aat aag cta aat ggt
1200
Cys Thr Ala Leu Asn Lys Phe Asn Val Tyr Gly Asn Lys Leu Asn Gly
385                 390                 395                 400

tct att cct gct ggt ttc cag aag ttg gag agt ctg act tac ttg aac
1248
Ser Ile Pro Ala Gly Phe Gln Lys Leu Glu Ser Leu Thr Tyr Leu Asn
                405                 410                 415

cta tct tca aac aat ttc aaa ggc aat att cct tct gag ctt ggt cac
1296
Leu Ser Ser Asn Asn Phe Lys Gly Asn Ile Pro Ser Glu Leu Gly His
            420                 425                 430

atc atc aac ttg gac aca ttg gat ctt tcc tac aat gaa ttc tct gga
1344
Ile Ile Asn Leu Asp Thr Leu Asp Leu Ser Tyr Asn Glu Phe Ser Gly
        435                 440                 445

cca gtt cct gct acc att ggt gat cta gag cac ctt ctt gaa ctg aat
1392
```

```
Pro Val Pro Ala Thr Ile Gly Asp Leu Glu His Leu Leu Glu Leu Asn
    450                 455                 460

ttg agt aag aac cat ctt gat ggg cca gtt cct gct gag ttt gga aac
1440
Leu Ser Lys Asn His Leu Asp Gly Pro Val Pro Ala Glu Phe Gly Asn
465                 470                 475                 480

ttg aga agc gtc caa gta att gat atg tcc aac aac aac tta tct ggt
1488
Leu Arg Ser Val Gln Val Ile Asp Met Ser Asn Asn Asn Leu Ser Gly
                485                 490                 495

agt ctg ccc gag gaa ctt gga caa ctt caa aac ctt gat agc ctg att
1536
Ser Leu Pro Glu Glu Leu Gly Gln Leu Gln Asn Leu Asp Ser Leu Ile
            500                 505                 510

ctt aac aac aac aat ttg gtt ggg gag atc cct gct caa ttg gcc aac
1584
Leu Asn Asn Asn Asn Leu Val Gly Glu Ile Pro Ala Gln Leu Ala Asn
        515                 520                 525

tgc ttc agc tta aat aac ctt gca ttt cag gaa ttt gtc ata caa caa
1632
Cys Phe Ser Leu Asn Asn Leu Ala Phe Gln Glu Phe Val Ile Gln Gln
    530                 535                 540

ttt atc tgg aca tgt ccc gat ggc aaa gaa ctt ctc gaa att ccc aat
1680
Phe Ile Trp Thr Cys Pro Asp Gly Lys Glu Leu Leu Glu Ile Pro Asn
545                 550                 555                 560

gga aag cat ctt cta att tct gat tgc aac cag tac ata aat cat aaa
1728
Gly Lys His Leu Leu Ile Ser Asp Cys Asn Gln Tyr Ile Asn His Lys
                565                 570                 575

tgc agc ttc ttg ggt aat cca tta ctg cat gtt tac tgc caa gat tcc
1776
Cys Ser Phe Leu Gly Asn Pro Leu Leu His Val Tyr Cys Gln Asp Ser
            580                 585                 590

agc tgt gga cac tct cat gga caa aga gtt aat att tca aag aca gca
1824
Ser Cys Gly His Ser His Gly Gln Arg Val Asn Ile Ser Lys Thr Ala
        595                 600                 605

att gct tgc att atc tta ggc ttt atc ata ttg ctc tgc gtt ctg ctg
1872
Ile Ala Cys Ile Ile Leu Gly Phe Ile Ile Leu Leu Cys Val Leu Leu
    610                 615                 620

ttg gct ata tat aaa aca aat caa cca cag cca ctt gtc aaa gga tcc
1920
Leu Ala Ile Tyr Lys Thr Asn Gln Pro Gln Pro Leu Val Lys Gly Ser
625                 630                 635                 640
```

```
gat aag cca gtg caa gga cct cca aag cta gtt gtt ctc cag atg gac
1968
Asp Lys Pro Val Gln Gly Pro Pro Lys Leu Val Val Leu Gln Met Asp
5                   645                 650                 655

atg gct atc cat act tac gag gac atc atg agg ctg aca gag aat ttg
2016
Met Ala Ile His Thr Tyr Glu Asp Ile Met Arg Leu Thr Glu Asn Leu
10              660                 665                 670

agc gag aaa tac atc att ggc tat ggc gcc tca agc act gtc tac aaa
2064
Ser Glu Lys Tyr Ile Ile Gly Tyr Gly Ala Ser Ser Thr Val Tyr Lys
15          675                 680                 685

tgt gaa ctc aag agc ggc aag gcc att gct gtc aag cgg ctt tac agt
2112
Cys Glu Leu Lys Ser Gly Lys Ala Ile Ala Val Lys Arg Leu Tyr Ser
20      690                 695                 700

cag tat aac cat agc ctc cga gag ttt gaa aca gaa cta gag aca att
2160
Gln Tyr Asn His Ser Leu Arg Glu Phe Glu Thr Glu Leu Glu Thr Ile
25  705                 710                 715                 720

ggc agc ata cgg cac agg aat ctt gtt agc ctc cat ggc ttc tcg cta
2208
Gly Ser Ile Arg His Arg Asn Leu Val Ser Leu His Gly Phe Ser Leu
30                  725                 730                 735

tct cca cat gga aac ttg ctc ttc tat gat tac atg gaa aat ggt tcc
2256
Ser Pro His Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Asn Gly Ser
35              740                 745                 750

ttg tgg gat ctt ctc cac ggt cca tca aag aaa gtg aag ctc aac tgg
2304
Leu Trp Asp Leu Leu His Gly Pro Ser Lys Lys Val Lys Leu Asn Trp
40          755                 760                 765

gac aca aga ctg agg atc gcg gtc gga gct gca caa ggg ctg gcc tat
2352
Asp Thr Arg Leu Arg Ile Ala Val Gly Ala Ala Gln Gly Leu Ala Tyr
45      770                 775                 780

ctc cac cat gac tgc aac cct cgc ata atc cac aga gat gtc aag tcc
2400
Leu His His Asp Cys Asn Pro Arg Ile Ile His Arg Asp Val Lys Ser
50  785                 790                 795                 800

tcc aac atc ctg ctc gac gag aac ttc gaa gcg cac ctc tca gat ttc
2448
Ser Asn Ile Leu Leu Asp Glu Asn Phe Glu Ala His Leu Ser Asp Phe
55                  805                 810                 815
```

```
ggc ata gcc aaa tgt gtc ccc tct gcc aag tcc cat gcc tcc act tat
2496
Gly Ile Ala Lys Cys Val Pro Ser Ala Lys Ser His Ala Ser Thr Tyr
            820                 825                 830

gtg cta gga acc atc ggc tac att gat ccg gag tat gcc agg act tcc
2544
Val Leu Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser
        835                 840                 845

agg ctc aat gag aaa tct gat gtg tac agc ttc ggc atc gtc ctt ctg
2592
Arg Leu Asn Glu Lys Ser Asp Val Tyr Ser Phe Gly Ile Val Leu Leu
    850                 855                 860

gaa ttg ctc aca ggg aag aag gcc gtc gac aac gaa tcg aac ttg cat
2640
Glu Leu Leu Thr Gly Lys Lys Ala Val Asp Asn Glu Ser Asn Leu His
865                 870                 875                 880

caa ttg ata ctc tcc aaa gct gat gac aac aca gtc atg gag gca gtg
2688
Gln Leu Ile Leu Ser Lys Ala Asp Asp Asn Thr Val Met Glu Ala Val
                885                 890                 895

gac tcg gag gtg tca gtg acg tgc acg gac atg gga ctg gtc agg aag
2736
Asp Ser Glu Val Ser Val Thr Cys Thr Asp Met Gly Leu Val Arg Lys
            900                 905                 910

gcc ttc cag ctc gcc ctt ctg tgc acc aag agg cac cct tca gac cgg
2784
Ala Phe Gln Leu Ala Leu Leu Cys Thr Lys Arg His Pro Ser Asp Arg
        915                 920                 925

ccg acc atg cac gag gtt gca agg gtg ctg ctc tcc ctg ctg ccg gcc
2832
Pro Thr Met His Glu Val Ala Arg Val Leu Leu Ser Leu Leu Pro Ala
    930                 935                 940

tcc gcc atg aca acg ccc aag acg gtg gac tac tcc cgg ttg ctg gcg
2880
Ser Ala Met Thr Thr Pro Lys Thr Val Asp Tyr Ser Arg Leu Leu Ala
945                 950                 955                 960

tcg acg acg acg gcg gcc gac atg cga ggg cac gac gtg acc gac atc
2928
Ser Thr Thr Thr Ala Ala Asp Met Arg Gly His Asp Val Thr Asp Ile
                965                 970                 975

ggc gac aac agc tcc tcc gac gag cag tgg ttc gtc agg ttc ggc gag
2976
Gly Asp Asn Ser Ser Ser Asp Glu Gln Trp Phe Val Arg Phe Gly Glu
            980                 985                 990

gtc ata tcc aag cac aca atg tga
3000
```

Val Ile Ser Lys His Thr Met
995

<210> 4

<211> 999

<212> PRT

<213> rice ERECTA homolog

<400> 4

Met Ala Ala Ala Arg Ala Pro Trp Leu Trp Trp Trp Val Val Val Val
1 5 10 15

Val Gly Val Ala Val Ala Glu Ala Ala Ser Gly Gly Gly Gly Gly Gly
20 25 30

Asp Gly Glu Gly Lys Ala Leu Met Gly Val Lys Ala Gly Phe Gly Asn
35 40 45

Ala Ala Asn Ala Leu Val Asp Trp Asp Gly Gly Ala Asp His Cys Ala
50 55 60

Trp Arg Gly Val Thr Cys Asp Asn Ala Ser Phe Ala Val Leu Ala Leu
65 70 75 80

Asn Leu Ser Asn Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly
85 90 95

Glu Leu Lys Asn Leu Gln Phe Val Asp Leu Lys Gly Asn Lys Leu Thr
100 105 110

Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys Ile Ser Leu Lys Tyr Leu
115 120 125

Asp Leu Ser Gly Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser
130 135 140

Lys Leu Lys Gln Leu Glu Glu Leu Ile Leu Lys Asn Asn Gln Leu Thr
145 150 155 160

Gly Pro Ile Pro Ser Thr Leu Ser Gln Ile Pro Asn Leu Lys Thr Leu

```
165                 170                 175

Asp Leu Ala Gln Asn Gln Leu Thr Gly Asp Ile Pro Arg Leu Ile Tyr
            180                 185                 190

Trp Asn Glu Val Leu Gln Tyr Leu Gly Leu Arg Gly Asn Ser Leu Thr
        195                 200                 205

Gly Thr Leu Ser Pro Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe
    210                 215                 220

Asp Val Arg Gly Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly
225                 230                 235                 240

Asn Cys Thr Ser Phe Glu Ile Leu Asp Ile Ser Tyr Asn Gln Ile Ser
                245                 250                 255

Gly Glu Ile Pro Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser
            260                 265                 270

Leu Gln Gly Asn Arg Leu Thr Gly Lys Ile Pro Asp Val Ile Gly Leu
        275                 280                 285

Met Gln Ala Leu Ala Val Leu Asp Leu Ser Glu Asn Glu Leu Val Gly
    290                 295                 300

Pro Ile Pro Ser Ile Leu Gly Asn Leu Ser Tyr Thr Gly Lys Leu Tyr
305                 310                 315                 320

Leu His Gly Asn Lys Leu Thr Gly Val Ile Pro Pro Glu Leu Gly Asn
                325                 330                 335

Met Ser Lys Leu Ser Tyr Leu Gln Leu Asn Asp Asn Glu Leu Val Gly
            340                 345                 350

Thr Ile Pro Ala Glu Leu Gly Lys Leu Glu Glu Leu Phe Glu Leu Asn
        355                 360                 365

Leu Ala Asn Asn Asn Leu Gln Gly Pro Ile Pro Ala Asn Ile Ser Ser
    370                 375                 380

Cys Thr Ala Leu Asn Lys Phe Asn Val Tyr Gly Asn Lys Leu Asn Gly
385                 390                 395                 400
```

Ser Ile Pro Ala Gly Phe Gln Lys Leu Glu Ser Leu Thr Tyr Leu Asn
405 410 415

Leu Ser Ser Asn Asn Phe Lys Gly Asn Ile Pro Ser Glu Leu Gly His
420 425 430

Ile Ile Asn Leu Asp Thr Leu Asp Leu Ser Tyr Asn Glu Phe Ser Gly
435 440 445

Pro Val Pro Ala Thr Ile Gly Asp Leu Glu His Leu Leu Glu Leu Asn
450 455 460

Leu Ser Lys Asn His Leu Asp Gly Pro Val Pro Ala Glu Phe Gly Asn
465 470 475 480

Leu Arg Ser Val Gln Val Ile Asp Met Ser Asn Asn Asn Leu Ser Gly
485 490 495

Ser Leu Pro Glu Glu Leu Gly Gln Leu Gln Asn Leu Asp Ser Leu Ile
500 505 510

Leu Asn Asn Asn Asn Leu Val Gly Glu Ile Pro Ala Gln Leu Ala Asn
515 520 525

Cys Phe Ser Leu Asn Asn Leu Ala Phe Gln Glu Phe Val Ile Gln Gln
530 535 540

Phe Ile Trp Thr Cys Pro Asp Gly Lys Glu Leu Leu Glu Ile Pro Asn
545 550 555 560

Gly Lys His Leu Leu Ile Ser Asp Cys Asn Gln Tyr Ile Asn His Lys
565 570 575

Cys Ser Phe Leu Gly Asn Pro Leu Leu His Val Tyr Cys Gln Asp Ser
580 585 590

Ser Cys Gly His Ser His Gly Gln Arg Val Asn Ile Ser Lys Thr Ala
595 600 605

Ile Ala Cys Ile Ile Leu Gly Phe Ile Ile Leu Leu Cys Val Leu Leu
610 615 620

Leu Ala Ile Tyr Lys Thr Asn Gln Pro Gln Pro Leu Val Lys Gly Ser
625 630 635 640

Asp Lys Pro Val Gln Gly Pro Pro Lys Leu Val Val Leu Gln Met Asp
645 650 655

Met Ala Ile His Thr Tyr Glu Asp Ile Met Arg Leu Thr Glu Asn Leu
660 665 670

Ser Glu Lys Tyr Ile Ile Gly Tyr Gly Ala Ser Ser Thr Val Tyr Lys
675 680 685

Cys Glu Leu Lys Ser Gly Lys Ala Ile Ala Val Lys Arg Leu Tyr Ser
690 695 700

Gln Tyr Asn His Ser Leu Arg Glu Phe Glu Thr Glu Leu Glu Thr Ile
705 710 715 720

Gly Ser Ile Arg His Arg Asn Leu Val Ser Leu His Gly Phe Ser Leu
725 730 735

Ser Pro His Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Asn Gly Ser
740 745 750

Leu Trp Asp Leu Leu His Gly Pro Ser Lys Lys Val Lys Leu Asn Trp
755 760 765

Asp Thr Arg Leu Arg Ile Ala Val Gly Ala Ala Gln Gly Leu Ala Tyr
770 775 780

Leu His His Asp Cys Asn Pro Arg Ile Ile His Arg Asp Val Lys Ser
785 790 795 800

Ser Asn Ile Leu Leu Asp Glu Asn Phe Glu Ala His Leu Ser Asp Phe
805 810 815

Gly Ile Ala Lys Cys Val Pro Ser Ala Lys Ser His Ala Ser Thr Tyr
820 825 830

Val Leu Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala Arg Thr Ser
835 840 845

Arg Leu Asn Glu Lys Ser Asp Val Tyr Ser Phe Gly Ile Val Leu Leu
850 855 860

Glu Leu Leu Thr Gly Lys Lys Ala Val Asp Asn Glu Ser Asn Leu His
865 870 875 880

Gln Leu Ile Leu Ser Lys Ala Asp Asp Asn Thr Val Met Glu Ala Val
885 890 895

Asp Ser Glu Val Ser Val Thr Cys Thr Asp Met Gly Leu Val Arg Lys
900 905 910

Ala Phe Gln Leu Ala Leu Leu Cys Thr Lys Arg His Pro Ser Asp Arg
915 920 925

Pro Thr Met His Glu Val Ala Arg Val Leu Leu Ser Leu Leu Pro Ala
930 935 940

Ser Ala Met Thr Thr Pro Lys Thr Val Asp Tyr Ser Arg Leu Leu Ala
945 950 955 960

Ser Thr Thr Thr Ala Ala Asp Met Arg Gly His Asp Val Thr Asp Ile
965 970 975

Gly Asp Asn Ser Ser Ser Asp Glu Gln Trp Phe Val Arg Phe Gly Glu
980 985 990

Val Ile Ser Lys His Thr Met
995

<210> 5

<211> 2766

<212> DNA

<213> sorghum ERECTA homolog

<220>

<221> CDS

<222> (1)..(2763)

<223>

<400> 5

```
atg acg acg acg gcc gcc cgt gct ctc gtc gcc ctc ctc ctc gtc gcc
Met Thr Thr Thr Ala Ala Arg Ala Leu Val Ala Leu Leu Leu Val Ala
1               5                   10                  15

gtc gcc gtc gcc gac gat ggg gcg acg ctg gtg gag atc aag aag tcc
Val Ala Val Ala Asp Asp Gly Ala Thr Leu Val Glu Ile Lys Lys Ser
            20                  25                  30

ttc cgc aac gtc ggc aac gta ctg tac gat tgg gcc ggc gac gac tac
Phe Arg Asn Val Gly Asn Val Leu Tyr Asp Trp Ala Gly Asp Asp Tyr
        35                  40                  45

tgc tcc tgg cgc ggc gtc ctg tgc gac aac gtc aca ttc gcc gtc gct
Cys Ser Trp Arg Gly Val Leu Cys Asp Asn Val Thr Phe Ala Val Ala
    50                  55                  60

gcg ctc aac ctc tct ggc ctc aac ctt gag ggc gag atc tct cca gcc
Ala Leu Asn Leu Ser Gly Leu Asn Leu Glu Gly Glu Ile Ser Pro Ala
65                  70                  75                  80

gtc ggc agc ctc aag agc ctc gtc tcc atc gat ctg aag tca aat ggg
Val Gly Ser Leu Lys Ser Leu Val Ser Ile Asp Leu Lys Ser Asn Gly
                85                  90                  95

cta tcc ggg cag atc cct gat gag att ggt gat tgt tca tca ctt agg
Leu Ser Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys Ser Ser Leu Arg
            100                 105                 110

acg ctg gac ttt tct ttc aac aac ttg gat ggc gac ata cca ttt tct
Thr Leu Asp Phe Ser Phe Asn Asn Leu Asp Gly Asp Ile Pro Phe Ser
        115                 120                 125

ata tca aag ctg aag cac ctg gag aac ttg ata ttg aag aac aac cag
Ile Ser Lys Leu Lys His Leu Glu Asn Leu Ile Leu Lys Asn Asn Gln
    130                 135                 140

ctg att ggt gcg atc cca tca aca ttg tca cag ctc cca aat ttg aag
Leu Ile Gly Ala Ile Pro Ser Thr Leu Ser Gln Leu Pro Asn Leu Lys
145                 150                 155                 160

att ttg gat ttg gca caa aac aaa ctg act ggg gag ata cca agg ctt
Ile Leu Asp Leu Ala Gln Asn Lys Leu Thr Gly Glu Ile Pro Arg Leu
```

```
    165                 170                 175

    atc tac tgg aat gag gtt ctt caa tat ctt gat gtg aag aac aat agc
    576
 5  Ile Tyr Trp Asn Glu Val Leu Gln Tyr Leu Asp Val Lys Asn Asn Ser
                180                 185                 190

    ttg acc ggg gtg ata cca gac acc att ggg aac tgt aca agt ttt caa
    624
10  Leu Thr Gly Val Ile Pro Asp Thr Ile Gly Asn Cys Thr Ser Phe Gln
            195                 200                 205

    gtc ttg gat ttg tct tac aac cgc ttt act gga cca atc cca ttc aac
    672
15  Val Leu Asp Leu Ser Tyr Asn Arg Phe Thr Gly Pro Ile Pro Phe Asn
        210                 215                 220

    att ggt ttc cta caa gtg gct aca cta tcc ttg caa ggg aac aag ttc
    720
20  Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Lys Phe
    225                 230                 235                 240

    acc ggt cca att cct tca gta att ggt ctt atg cag gct ctc gct gtt
    768
25  Thr Gly Pro Ile Pro Ser Val Ile Gly Leu Met Gln Ala Leu Ala Val
                    245                 250                 255

    cta gat ctg agt tac aac caa tta tct ggt cct ata cca tca ata cta
    816
30  Leu Asp Leu Ser Tyr Asn Gln Leu Ser Gly Pro Ile Pro Ser Ile Leu
                260                 265                 270

    ggc aac ttg aca tac act gag aag ctg tac atc caa ggc aat aag tta
    864
35  Gly Asn Leu Thr Tyr Thr Glu Lys Leu Tyr Ile Gln Gly Asn Lys Leu
            275                 280                 285

    act ggg tcg ata cca cca gag tta gga aat atg tca aca ctt cat tac
    912
40  Thr Gly Ser Ile Pro Pro Glu Leu Gly Asn Met Ser Thr Leu His Tyr
        290                 295                 300

    cta gaa ctg aac gat aat caa ctt act ggg tca att cca cca gag ctt
    960
45  Leu Glu Leu Asn Asp Asn Gln Leu Thr Gly Ser Ile Pro Pro Glu Leu
    305                 310                 315                 320

    gga agg cta aca ggc ttg ttt gac ctg aac ctt gcg aat aac cac ctg
    1008
50  Gly Arg Leu Thr Gly Leu Phe Asp Leu Asn Leu Ala Asn Asn His Leu
                    325                 330                 335

    gaa gga cca att cct gac aac cta agt tca tgt gtg aat ctc aat agc
    1056
55  Glu Gly Pro Ile Pro Asp Asn Leu Ser Ser Cys Val Asn Leu Asn Ser
                340                 345                 350
```

```
ttc aat gct tat ggc aac aag tta aat ggg acc att cct cgt tcg ttg
1104
Phe Asn Ala Tyr Gly Asn Lys Leu Asn Gly Thr Ile Pro Arg Ser Leu
        355                 360                 365

cgg aaa ctt gaa agc atg acc tat tta aat ctg tca tca aac ttc ata
1152
Arg Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu Ser Ser Asn Phe Ile
    370                 375                 380

agt ggc tct att cct att gag tta tca agg atc aac aat ttg gac acg
1200
Ser Gly Ser Ile Pro Ile Glu Leu Ser Arg Ile Asn Asn Leu Asp Thr
385                 390                 395                 400

ctg gat tta tcc tgt aac atg atg act ggt cca att cca tca tca att
1248
Leu Asp Leu Ser Cys Asn Met Met Thr Gly Pro Ile Pro Ser Ser Ile
                405                 410                 415

ggc agc cta gag cat cta ttg aga ctt aac ttg agc aag aat ggt cta
1296
Gly Ser Leu Glu His Leu Leu Arg Leu Asn Leu Ser Lys Asn Gly Leu
            420                 425                 430

gtt gga ttc atc ccc gcg gag ttt ggt aat ttg agg agt gtc atg gag
1344
Val Gly Phe Ile Pro Ala Glu Phe Gly Asn Leu Arg Ser Val Met Glu
        435                 440                 445

att gat tta tcc tat aat cac ctt ggt ggc ctg att cct caa gaa ctt
1392
Ile Asp Leu Ser Tyr Asn His Leu Gly Gly Leu Ile Pro Gln Glu Leu
    450                 455                 460

gaa atg ctg caa aac ctg atg ttg cta aat gtg tcg tac aat aat ttg
1440
Glu Met Leu Gln Asn Leu Met Leu Leu Asn Val Ser Tyr Asn Asn Leu
465                 470                 475                 480

gct ggt gtt gtc cct gct gac aac aac ttc aca cgg ttt tca cct gac
1488
Ala Gly Val Val Pro Ala Asp Asn Asn Phe Thr Arg Phe Ser Pro Asp
                485                 490                 495

agc ttt tta ggt aat cct gga ctc tgt gga tac tgg ctt ggt tcg tcg
1536
Ser Phe Leu Gly Asn Pro Gly Leu Cys Gly Tyr Trp Leu Gly Ser Ser
            500                 505                 510

tgt cgt tcc act ggc cac cac gag aaa ccg cct atc tca aag gct gcc
1584
Cys Arg Ser Thr Gly His His Glu Lys Pro Pro Ile Ser Lys Ala Ala
        515                 520                 525

ata att ggt gtt gct gtg ggt gga ctt gtt atc ctc ttg atg atc tta
1632
```

```
Ile Ile Gly Val Ala Val Gly Gly Leu Val Ile Leu Leu Met Ile Leu
    530                 535                 540

gta gct gtt tgc agg cca cat cgt cca cct gct ttt aaa gat gtc act
1680
Val Ala Val Cys Arg Pro His Arg Pro Pro Ala Phe Lys Asp Val Thr
545                 550                 555                 560

gta agc aag cca gtg aga aat gct ccc ccc aag ctg gtg atc ctt cat
1728
Val Ser Lys Pro Val Arg Asn Ala Pro Pro Lys Leu Val Ile Leu His
                565                 570                 575

atg aac atg gcc ctt cat gta tac gat gac ata atg agg atg act gag
1776
Met Asn Met Ala Leu His Val Tyr Asp Asp Ile Met Arg Met Thr Glu
            580                 585                 590

aac ttg agt gag aaa tac atc att gga tac ggg gcg tca agt aca gtt
1824
Asn Leu Ser Glu Lys Tyr Ile Ile Gly Tyr Gly Ala Ser Ser Thr Val
        595                 600                 605

tat aaa tgt gtc cta aag aat tgc aaa ccg gtg gca ata aaa aag ctg
1872
Tyr Lys Cys Val Leu Lys Asn Cys Lys Pro Val Ala Ile Lys Lys Leu
    610                 615                 620

tat gcc cac tac cca cag agc ctt aag gaa ttt gaa act gag ctt gag
1920
Tyr Ala His Tyr Pro Gln Ser Leu Lys Glu Phe Glu Thr Glu Leu Glu
625                 630                 635                 640

act gtt ggt agc atc aag cac cgg aat cta gtc agc ctt caa ggg tac
1968
Thr Val Gly Ser Ile Lys His Arg Asn Leu Val Ser Leu Gln Gly Tyr
                645                 650                 655

tca tta tca cct gtt ggg aac ctc ctc ttt tat gat tat atg gaa tgt
2016
Ser Leu Ser Pro Val Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Cys
            660                 665                 670

ggc agc tta tgg gat gtt tta cat gaa ggt tca tcc aag aag aaa aaa
2064
Gly Ser Leu Trp Asp Val Leu His Glu Gly Ser Ser Lys Lys Lys Lys
        675                 680                 685

ctt gac tgg gag act cgc cta cgg att gct ctt ggt gca gct caa ggc
2112
Leu Asp Trp Glu Thr Arg Leu Arg Ile Ala Leu Gly Ala Ala Gln Gly
    690                 695                 700

ctt gct tac ctt cac cat gac tgc agt cca cgg ata att cat cgg gat
2160
Leu Ala Tyr Leu His His Asp Cys Ser Pro Arg Ile Ile His Arg Asp
705                 710                 715                 720
```

```
gta aaa tca aag aat ata ctc ctt gac aaa gat tat gag gcc cat ctt
2208
Val Lys Ser Lys Asn Ile Leu Leu Asp Lys Asp Tyr Glu Ala His Leu
                725                 730                 735

aca gac ttt gga att gct aag agc tta tgt gtc tca aaa act cac aca
2256
Thr Asp Phe Gly Ile Ala Lys Ser Leu Cys Val Ser Lys Thr His Thr
            740                 745                 750

tca acc tat gtc atg gga act att ggc tac att gat cct gag tac gcc
2304
Ser Thr Tyr Val Met Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala
        755                 760                 765

cgc act tcc cgt ctc aac gaa aag tct gat gtc tac agg cta tgg cat
2352
Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Val Tyr Arg Leu Trp His
    770                 775                 780

tgt tct gct gga gct gct gac tgg caa gaa gcc agt gga caa cga atc
2400
Cys Ser Ala Gly Ala Ala Asp Trp Gln Glu Ala Ser Gly Gln Arg Ile
785                 790                 795                 800

cta tcg aag acg gca agc aac gag gtc atg gat acc gtg gac cct gac
2448
Leu Ser Lys Thr Ala Ser Asn Glu Val Met Asp Thr Val Asp Pro Asp
                805                 810                 815

atc ggg gac acc tgc aag gac ctc ggc gag gtg aag aag ctg ttc cag
2496
Ile Gly Asp Thr Cys Lys Asp Leu Gly Glu Val Lys Lys Leu Phe Gln
            820                 825                 830

ctg gcg ctc ctt tgc acc aag cgg caa ccc tcg gac cga ccg acg atg
2544
Leu Ala Leu Leu Cys Thr Lys Arg Gln Pro Ser Asp Arg Pro Thr Met
        835                 840                 845

cac gag gtg gtg cgc gtc ctg gac tgc ctg gtg aac ccg gac ccg ccg
2592
His Glu Val Val Arg Val Leu Asp Cys Leu Val Asn Pro Asp Pro Pro
    850                 855                 860

cca aag ccg tcg gcg cac cag ctg ccg cag ccg tcg cca gcc gtg cca
2640
Pro Lys Pro Ser Ala His Gln Leu Pro Gln Pro Ser Pro Ala Val Pro
865                 870                 875                 880

agc tac atc aac gag tac gtc agc ctg cgg ggc acc ggc gct ctc tcc
2688
Ser Tyr Ile Asn Glu Tyr Val Ser Leu Arg Gly Thr Gly Ala Leu Ser
                885                 890                 895
```

```
tgc gcc aac tcg acc agc acc tcg gac gcc gag ctg ttc ctc aag ttc
2736
Cys Ala Asn Ser Thr Ser Thr Ser Asp Ala Glu Leu Phe Leu Lys Phe
             900                 905                 910

ggc gag gcc atc tcg cag aac atg gag tag
2766
Gly Glu Ala Ile Ser Gln Asn Met Glu
        915                 920
```

<210> 6

<211> 921

<212> PRT

<213> sorghum ERECTA homolog

<400> 6

```
Met Thr Thr Thr Ala Ala Arg Ala Leu Val Ala Leu Leu Leu Val Ala
1               5                   10                  15

Val Ala Val Ala Asp Asp Gly Ala Thr Leu Val Glu Ile Lys Lys Ser
            20                  25                  30

Phe Arg Asn Val Gly Asn Val Leu Tyr Asp Trp Ala Gly Asp Asp Tyr
        35                  40                  45

Cys Ser Trp Arg Gly Val Leu Cys Asp Asn Val Thr Phe Ala Val Ala
    50                  55                  60

Ala Leu Asn Leu Ser Gly Leu Asn Leu Glu Gly Glu Ile Ser Pro Ala
65                  70                  75                  80

Val Gly Ser Leu Lys Ser Leu Val Ser Ile Asp Leu Lys Ser Asn Gly
                85                  90                  95

Leu Ser Gly Gln Ile Pro Asp Glu Ile Gly Asp Cys Ser Ser Leu Arg
            100                 105                 110

Thr Leu Asp Phe Ser Phe Asn Asn Leu Asp Gly Asp Ile Pro Phe Ser
        115                 120                 125

Ile Ser Lys Leu Lys His Leu Glu Asn Leu Ile Leu Lys Asn Asn Gln
    130                 135                 140
```

```
Leu Ile Gly Ala Ile Pro Ser Thr Leu Ser Gln Leu Pro Asn Leu Lys
145                 150                 155                 160

Ile Leu Asp Leu Ala Gln Asn Lys Leu Thr Gly Glu Ile Pro Arg Leu
                165                 170                 175

Ile Tyr Trp Asn Glu Val Leu Gln Tyr Leu Asp Val Lys Asn Asn Ser
            180                 185                 190

Leu Thr Gly Val Ile Pro Asp Thr Ile Gly Asn Cys Thr Ser Phe Gln
        195                 200                 205

Val Leu Asp Leu Ser Tyr Asn Arg Phe Thr Gly Pro Ile Pro Phe Asn
    210                 215                 220

Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Lys Phe
225                 230                 235                 240

Thr Gly Pro Ile Pro Ser Val Ile Gly Leu Met Gln Ala Leu Ala Val
                245                 250                 255

Leu Asp Leu Ser Tyr Asn Gln Leu Ser Gly Pro Ile Pro Ser Ile Leu
            260                 265                 270

Gly Asn Leu Thr Tyr Thr Glu Lys Leu Tyr Ile Gln Gly Asn Lys Leu
        275                 280                 285

Thr Gly Ser Ile Pro Pro Glu Leu Gly Asn Met Ser Thr Leu His Tyr
    290                 295                 300

Leu Glu Leu Asn Asp Asn Gln Leu Thr Gly Ser Ile Pro Pro Glu Leu
305                 310                 315                 320

Gly Arg Leu Thr Gly Leu Phe Asp Leu Asn Leu Ala Asn Asn His Leu
                325                 330                 335

Glu Gly Pro Ile Pro Asp Asn Leu Ser Ser Cys Val Asn Leu Asn Ser
            340                 345                 350

Phe Asn Ala Tyr Gly Asn Lys Leu Asn Gly Thr Ile Pro Arg Ser Leu
        355                 360                 365
```

```
Arg Lys Leu Glu Ser Met Thr Tyr Leu Asn Leu Ser Ser Asn Phe Ile
370                 375                 380

Ser Gly Ser Ile Pro Ile Glu Leu Ser Arg Ile Asn Asn Leu Asp Thr
385                 390                 395                 400

Leu Asp Leu Ser Cys Asn Met Met Thr Gly Pro Ile Pro Ser Ser Ile
                405                 410                 415

Gly Ser Leu Glu His Leu Leu Arg Leu Asn Leu Ser Lys Asn Gly Leu
            420                 425                 430

Val Gly Phe Ile Pro Ala Glu Phe Gly Asn Leu Arg Ser Val Met Glu
        435                 440                 445

Ile Asp Leu Ser Tyr Asn His Leu Gly Gly Leu Ile Pro Gln Glu Leu
    450                 455                 460

Glu Met Leu Gln Asn Leu Met Leu Leu Asn Val Ser Tyr Asn Asn Leu
465                 470                 475                 480

Ala Gly Val Val Pro Ala Asp Asn Asn Phe Thr Arg Phe Ser Pro Asp
                485                 490                 495

Ser Phe Leu Gly Asn Pro Gly Leu Cys Gly Tyr Trp Leu Gly Ser Ser
            500                 505                 510

Cys Arg Ser Thr Gly His His Glu Lys Pro Pro Ile Ser Lys Ala Ala
        515                 520                 525

Ile Ile Gly Val Ala Val Gly Gly Leu Val Ile Leu Leu Met Ile Leu
    530                 535                 540

Val Ala Val Cys Arg Pro His Arg Pro Pro Ala Phe Lys Asp Val Thr
545                 550                 555                 560

Val Ser Lys Pro Val Arg Asn Ala Pro Pro Lys Leu Val Ile Leu His
                565                 570                 575

Met Asn Met Ala Leu His Val Tyr Asp Asp Ile Met Arg Met Thr Glu
            580                 585                 590
```

Asn Leu Ser Glu Lys Tyr Ile Ile Gly Tyr Gly Ala Ser Ser Thr Val
595 600 605

Tyr Lys Cys Val Leu Lys Asn Cys Lys Pro Val Ala Ile Lys Lys Leu
610 615 620

Tyr Ala His Tyr Pro Gln Ser Leu Lys Glu Phe Glu Thr Glu Leu Glu
625 630 635 640

Thr Val Gly Ser Ile Lys His Arg Asn Leu Val Ser Leu Gln Gly Tyr
645 650 655

Ser Leu Ser Pro Val Gly Asn Leu Leu Phe Tyr Asp Tyr Met Glu Cys
660 665 670

Gly Ser Leu Trp Asp Val Leu His Glu Gly Ser Ser Lys Lys Lys Lys
675 680 685

Leu Asp Trp Glu Thr Arg Leu Arg Ile Ala Leu Gly Ala Ala Gln Gly
690 695 700

Leu Ala Tyr Leu His His Asp Cys Ser Pro Arg Ile Ile His Arg Asp
705 710 715 720

Val Lys Ser Lys Asn Ile Leu Leu Asp Lys Asp Tyr Glu Ala His Leu
725 730 735

Thr Asp Phe Gly Ile Ala Lys Ser Leu Cys Val Ser Lys Thr His Thr
740 745 750

Ser Thr Tyr Val Met Gly Thr Ile Gly Tyr Ile Asp Pro Glu Tyr Ala
755 760 765

Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Val Tyr Arg Leu Trp His
770 775 780

Cys Ser Ala Gly Ala Ala Asp Trp Gln Glu Ala Ser Gly Gln Arg Ile
785 790 795 800

Leu Ser Lys Thr Ala Ser Asn Glu Val Met Asp Thr Val Asp Pro Asp
805 810 815

Ile Gly Asp Thr Cys Lys Asp Leu Gly Glu Val Lys Lys Leu Phe Gln

820 825 830

Leu Ala Leu Leu Cys Thr Lys Arg Gln Pro Ser Asp Arg Pro Thr Met
835 840 845

His Glu Val Val Arg Val Leu Asp Cys Leu Val Asn Pro Asp Pro Pro
850 855 860

Pro Lys Pro Ser Ala His Gln Leu Pro Gln Pro Ser Pro Ala Val Pro
865 870 875 880

Ser Tyr Ile Asn Glu Tyr Val Ser Leu Arg Gly Thr Gly Ala Leu Ser
885 890 895

Cys Ala Asn Ser Thr Ser Thr Ser Asp Ala Glu Leu Phe Leu Lys Phe
900 905 910

Gly Glu Ala Ile Ser Gln Asn Met Glu
915 920

<210> 7

<211> 2799

<212> DNA

<213> Arabdopsis thaliana ERECTA homolog

<220>

<221> CDS

<222> (1)..(2796)

<223>

<400> 7

atg gcg ata aag gct tca ttc agc aac gtg gcg aat atg ctt ctt gat
48
Met Ala Ile Lys Ala Ser Phe Ser Asn Val Ala Asn Met Leu Leu Asp
1 5 10 15

tgg gac gat gtt cat aac cac gac ttt tgt tct tgg aga ggt gtc ttc
96
Trp Asp Asp Val His Asn His Asp Phe Cys Ser Trp Arg Gly Val Phe
20 25 30

```
tgt gat aac gtt agc ctc aat gtt gtc tct ctt aat ctg tca aac ctg
144
Cys Asp Asn Val Ser Leu Asn Val Val Ser Leu Asn Leu Ser Asn Leu
        35                  40                  45

aat ctt ggt gga gag ata tca tct gcc ctt gga gat ttg atg aat ctg
192
Asn Leu Gly Gly Glu Ile Ser Ser Ala Leu Gly Asp Leu Met Asn Leu
    50                  55                  60

caa tca ata gac ttg caa gga aat aaa ttg ggt ggt caa att cca gat
240
Gln Ser Ile Asp Leu Gln Gly Asn Lys Leu Gly Gly Gln Ile Pro Asp
65                  70                  75                  80

gag att gga aac tgt gtt tct ctt gct tat gtg gat ttc tcc acc aat
288
Glu Ile Gly Asn Cys Val Ser Leu Ala Tyr Val Asp Phe Ser Thr Asn
                85                  90                  95

ttg ttg ttt gga gac ata ccg ttt tca atc tct aaa ctc aaa cag ctg
336
Leu Leu Phe Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln Leu
            100                 105                 110

gag ttt ctg aac cta aag aat aat cag ctc act ggt cca ata cca gca
384
Glu Phe Leu Asn Leu Lys Asn Asn Gln Leu Thr Gly Pro Ile Pro Ala
        115                 120                 125

acc tta act cag att cca aac ctt aag acc ctt gac ctc gca aga aac
432
Thr Leu Thr Gln Ile Pro Asn Leu Lys Thr Leu Asp Leu Ala Arg Asn
    130                 135                 140

cag ctt act ggt gag ata cca agg tta ctc tac tgg aat gaa gtt tta
480
Gln Leu Thr Gly Glu Ile Pro Arg Leu Leu Tyr Trp Asn Glu Val Leu
145                 150                 155                 160

cag tat ctc ggt tta cgt ggg aat atg tta act ggg aca ttg tct cct
528
Gln Tyr Leu Gly Leu Arg Gly Asn Met Leu Thr Gly Thr Leu Ser Pro
                165                 170                 175

gat atg tgt cag ctg acg ggt ctg tgg tac ttt gat gtg aga ggc aac
576
Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly Asn
            180                 185                 190

aac ctt act gga act atc cca gag agc att ggc aat tgc aca agc ttt
624
Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly Asn Cys Thr Ser Phe
        195                 200                 205
```

```
gag atc ttg gat gta tct tat aat cag att acc gga gtt ata ccc tac
672
Glu Ile Leu Asp Val Ser Tyr Asn Gln Ile Thr Gly Val Ile Pro Tyr
    210                 215                 220

aat att ggt ttc ctc caa gta gct act ctg tca ctt caa gga aac aag
720
Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Lys
225                 230                 235                 240

ttg act ggc aga att ccg gaa gtg att ggt ctg atg cag gct ctt gct
768
Leu Thr Gly Arg Ile Pro Glu Val Ile Gly Leu Met Gln Ala Leu Ala
                245                 250                 255

gta ttg gat ttg agt gac aat gaa tta act ggg cct att cca cca ata
816
Val Leu Asp Leu Ser Asp Asn Glu Leu Thr Gly Pro Ile Pro Pro Ile
            260                 265                 270

ctt ggg aat ctg tca ttc act gga aaa ctg tat ctc cat ggc aac aag
864
Leu Gly Asn Leu Ser Phe Thr Gly Lys Leu Tyr Leu His Gly Asn Lys
        275                 280                 285

ctc act gga caa atc cca ccc gag cta ggc aat atg tca cga ctc agc
912
Leu Thr Gly Gln Ile Pro Pro Glu Leu Gly Asn Met Ser Arg Leu Ser
    290                 295                 300

tat ttg caa cta aat gat aat gaa cta gtg gga aag atc cca cct gag
960
Tyr Leu Gln Leu Asn Asp Asn Glu Leu Val Gly Lys Ile Pro Pro Glu
305                 310                 315                 320

ctt ggg aag ctg gaa caa ttg ttc gaa ctg aat ctt gcg aac aac aat
1008
Leu Gly Lys Leu Glu Gln Leu Phe Glu Leu Asn Leu Ala Asn Asn Asn
                325                 330                 335

ctt gta ggg ctg att cca tct aac att agt tcc tgt gct gcc ttg aat
1056
Leu Val Gly Leu Ile Pro Ser Asn Ile Ser Ser Cys Ala Ala Leu Asn
            340                 345                 350

caa ttc aat gtt cat ggg aac ttc ttg agt gga gct gta cca ctt gaa
1104
Gln Phe Asn Val His Gly Asn Phe Leu Ser Gly Ala Val Pro Leu Glu
        355                 360                 365

ttc cgg aat ctt gga agc ttg act tat cta aat ctt tcc tca aac agt
1152
Phe Arg Asn Leu Gly Ser Leu Thr Tyr Leu Asn Leu Ser Ser Asn Ser
    370                 375                 380

ttc aag ggc aaa ata cct gct gag ctt ggc cat atc atc aat ctt gat
1200
```

```
Phe Lys Gly Lys Ile Pro Ala Glu Leu Gly His Ile Ile Asn Leu Asp
385                 390                 395                 400

aca ttg gat ctg tct ggc aac aat ttc tca ggc tca att cca tta aca
1248
Thr Leu Asp Leu Ser Gly Asn Asn Phe Ser Gly Ser Ile Pro Leu Thr
                405                 410                 415

ctt ggt gat ctt gag cat ctt ctc atc tta aac ttg agc aga aat cat
1296
Leu Gly Asp Leu Glu His Leu Leu Ile Leu Asn Leu Ser Arg Asn His
            420                 425                 430

ctg aat ggc aca ttg cct gca gaa ttc ggg aac ctc cga agc att cag
1344
Leu Asn Gly Thr Leu Pro Ala Glu Phe Gly Asn Leu Arg Ser Ile Gln
        435                 440                 445

atc atc gat gtg tca ttt aat ttt ctt gcc ggt gtt att cca act gaa
1392
Ile Ile Asp Val Ser Phe Asn Phe Leu Ala Gly Val Ile Pro Thr Glu
    450                 455                 460

ctt ggc cag ttg cag aac ata aac tct ctg ata ctg aac aac aac aag
1440
Leu Gly Gln Leu Gln Asn Ile Asn Ser Leu Ile Leu Asn Asn Asn Lys
465                 470                 475                 480

att cat ggg aaa atc cct gat cag cta act aac tgc ttc agt ctt gcc
1488
Ile His Gly Lys Ile Pro Asp Gln Leu Thr Asn Cys Phe Ser Leu Ala
                485                 490                 495

aat ctg aac atc tcc ttc aat aat ctt tct gga ata atc cca cct atg
1536
Asn Leu Asn Ile Ser Phe Asn Asn Leu Ser Gly Ile Ile Pro Pro Met
            500                 505                 510

aag aac ttt aca cgt ttt tcc ccg gcc agc ttc ttt gga aat cca ttt
1584
Lys Asn Phe Thr Arg Phe Ser Pro Ala Ser Phe Phe Gly Asn Pro Phe
        515                 520                 525

ctc tgc ggg aac tgg gtt gga tca atc tgt ggc cca tct tta cct aag
1632
Leu Cys Gly Asn Trp Val Gly Ser Ile Cys Gly Pro Ser Leu Pro Lys
    530                 535                 540

tca caa gta ttc acc aga gtt gcc gtg att tgt atg gtt ctc ggt ttc
1680
Ser Gln Val Phe Thr Arg Val Ala Val Ile Cys Met Val Leu Gly Phe
545                 550                 555                 560

atc act ctc ata tgc atg ata ttc att gcg gtt tac aag tca aag cag
1728
Ile Thr Leu Ile Cys Met Ile Phe Ile Ala Val Tyr Lys Ser Lys Gln
                565                 570                 575
```

```
cag aaa cca gtc ttg aaa ggc tct tca aaa caa cct gaa ggg tca acg
1776
Gln Lys Pro Val Leu Lys Gly Ser Ser Lys Gln Pro Glu Gly Ser Thr
            580                 585                 590

aag ctg gtg att ctt cac atg gac atg gct att cac acg ttt gat gat
1824
Lys Leu Val Ile Leu His Met Asp Met Ala Ile His Thr Phe Asp Asp
        595                 600                 605

atc atg aga gtt aca gaa aac ctc gat gag aaa tac atc att gga tac
1872
Ile Met Arg Val Thr Glu Asn Leu Asp Glu Lys Tyr Ile Ile Gly Tyr
    610                 615                 620

ggt gct tct agc aca gtt tac aag tgc acc tcc aaa act tcc cga cct
1920
Gly Ala Ser Ser Thr Val Tyr Lys Cys Thr Ser Lys Thr Ser Arg Pro
625                 630                 635                 640

att gcc att aag cga atc tac aat cag tat ccc agc aac ttc cgc gag
1968
Ile Ala Ile Lys Arg Ile Tyr Asn Gln Tyr Pro Ser Asn Phe Arg Glu
                645                 650                 655

ttt gaa aca gag ctc gag acc att ggg agc atc aga cac aga aac ata
2016
Phe Glu Thr Glu Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Ile
            660                 665                 670

gta agc ttg cac gga tac gcc tta tct ccc ttt ggc aac ctc ctc ttc
2064
Val Ser Leu His Gly Tyr Ala Leu Ser Pro Phe Gly Asn Leu Leu Phe
        675                 680                 685

tac gac tac atg gaa aat ggc tct ctt tgg gat ctt ctc cat ggg cct
2112
Tyr Asp Tyr Met Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro
    690                 695                 700

ggg aag aag gtg aag ctt gac tgg gaa aca agg ctg aag ata gct gtt
2160
Gly Lys Lys Val Lys Leu Asp Trp Glu Thr Arg Leu Lys Ile Ala Val
705                 710                 715                 720

gga gct gcg caa gga ctt gca tat ctt cac cat gac tgc aca cct agg
2208
Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Thr Pro Arg
                725                 730                 735

ata atc cat cga gac atc aag tca tca aac ata ctc ctt gat ggg aat
2256
Ile Ile His Arg Asp Ile Lys Ser Ser Asn Ile Leu Leu Asp Gly Asn
            740                 745                 750
```

```
ttc gaa gcg cgt ttg tca gat ttt ggg att gcc aag agc ata cca gcc
2304
Phe Glu Ala Arg Leu Ser Asp Phe Gly Ile Ala Lys Ser Ile Pro Ala
        755                 760                 765

acc aaa act tat gct tca acc tat gtt ctt gga acc att gga tat att
2352
Thr Lys Thr Tyr Ala Ser Thr Tyr Val Leu Gly Thr Ile Gly Tyr Ile
    770                 775                 780

gac cca gag tat gct cga act tcg cgt ctg aac gag aag tct gat atc
2400
Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Ile
785                 790                 795                 800

tac agt ttc ggt att gtc ctt ctt gag ctt cta acc ggc aag aag gct
2448
Tyr Ser Phe Gly Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Ala
                805                 810                 815

gtg gat aac gag gcc aac ttg cat caa atg att cta tca aag gcg gat
2496
Val Asp Asn Glu Ala Asn Leu His Gln Met Ile Leu Ser Lys Ala Asp
            820                 825                 830

gat aac aca gta atg gaa gct gtt gat gca gag gtc tca gtg act tgc
2544
Asp Asn Thr Val Met Glu Ala Val Asp Ala Glu Val Ser Val Thr Cys
        835                 840                 845

atg gac tca gga cac atc aag aaa aca ttt cag cta gct ctc ttg tgc
2592
Met Asp Ser Gly His Ile Lys Lys Thr Phe Gln Leu Ala Leu Leu Cys
    850                 855                 860

acc aag cga aat cct ttg gag aga ccc acc atg cag gag gtc tct agg
2640
Thr Lys Arg Asn Pro Leu Glu Arg Pro Thr Met Gln Glu Val Ser Arg
865                 870                 875                 880

gtt ctg ctc tca ctt gtc ccg tct cca cct cca aag aag tta ccg tcg
2688
Val Leu Leu Ser Leu Val Pro Ser Pro Pro Pro Lys Lys Leu Pro Ser
                885                 890                 895

cct gca aaa gta cag gaa ggg gaa gaa cgg cgt gag agc cac tct tca
2736
Pro Ala Lys Val Gln Glu Gly Glu Glu Arg Arg Glu Ser His Ser Ser
            900                 905                 910

gat aca aca acc cca cag tgg ttt gtt cag ttc cgt gaa gat atc tcc
2784
Asp Thr Thr Thr Pro Gln Trp Phe Val Gln Phe Arg Glu Asp Ile Ser
        915                 920                 925

aaa agt agc tta taa
2799
```

Lys Ser Ser Leu
930

<210> 8

<211> 932

<212> PRT

<213> Arabdopsis thaliana ERECTA homolog

<400> 8

Met Ala Ile Lys Ala Ser Phe Ser Asn Val Ala Asn Met Leu Leu Asp
1 5 10 15

Trp Asp Asp Val His Asn His Asp Phe Cys Ser Trp Arg Gly Val Phe
20 25 30

Cys Asp Asn Val Ser Leu Asn Val Val Ser Leu Asn Leu Ser Asn Leu
35 40 45

Asn Leu Gly Gly Glu Ile Ser Ser Ala Leu Gly Asp Leu Met Asn Leu
50 55 60

Gln Ser Ile Asp Leu Gln Gly Asn Lys Leu Gly Gly Gln Ile Pro Asp
65 70 75 80

Glu Ile Gly Asn Cys Val Ser Leu Ala Tyr Val Asp Phe Ser Thr Asn
85 90 95

Leu Leu Phe Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln Leu
100 105 110

Glu Phe Leu Asn Leu Lys Asn Asn Gln Leu Thr Gly Pro Ile Pro Ala
115 120 125

Thr Leu Thr Gln Ile Pro Asn Leu Lys Thr Leu Asp Leu Ala Arg Asn
130 135 140

Gln Leu Thr Gly Glu Ile Pro Arg Leu Leu Tyr Trp Asn Glu Val Leu
145 150 155 160

Gln Tyr Leu Gly Leu Arg Gly Asn Met Leu Thr Gly Thr Leu Ser Pro

```
165                 170                 175

Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly Asn
            180                 185                 190

Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly Asn Cys Thr Ser Phe
        195                 200                 205

Glu Ile Leu Asp Val Ser Tyr Asn Gln Ile Thr Gly Val Ile Pro Tyr
    210                 215                 220

Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn Lys
225                 230                 235                 240

Leu Thr Gly Arg Ile Pro Glu Val Ile Gly Leu Met Gln Ala Leu Ala
                245                 250                 255

Val Leu Asp Leu Ser Asp Asn Glu Leu Thr Gly Pro Ile Pro Pro Ile
            260                 265                 270

Leu Gly Asn Leu Ser Phe Thr Gly Lys Leu Tyr Leu His Gly Asn Lys
        275                 280                 285

Leu Thr Gly Gln Ile Pro Pro Glu Leu Gly Asn Met Ser Arg Leu Ser
    290                 295                 300

Tyr Leu Gln Leu Asn Asp Asn Glu Leu Val Gly Lys Ile Pro Pro Glu
305                 310                 315                 320

Leu Gly Lys Leu Glu Gln Leu Phe Glu Leu Asn Leu Ala Asn Asn Asn
                325                 330                 335

Leu Val Gly Leu Ile Pro Ser Asn Ile Ser Ser Cys Ala Ala Leu Asn
            340                 345                 350

Gln Phe Asn Val His Gly Asn Phe Leu Ser Gly Ala Val Pro Leu Glu
        355                 360                 365

Phe Arg Asn Leu Gly Ser Leu Thr Tyr Leu Asn Leu Ser Ser Asn Ser
    370                 375                 380

Phe Lys Gly Lys Ile Pro Ala Glu Leu Gly His Ile Ile Asn Leu Asp
385                 390                 395                 400
```

Thr Leu Asp Leu Ser Gly Asn Asn Phe Ser Gly Ser Ile Pro Leu Thr
405 410 415

Leu Gly Asp Leu Glu His Leu Leu Ile Leu Asn Leu Ser Arg Asn His
420 425 430

Leu Asn Gly Thr Leu Pro Ala Glu Phe Gly Asn Leu Arg Ser Ile Gln
435 440 445

Ile Ile Asp Val Ser Phe Asn Phe Leu Ala Gly Val Ile Pro Thr Glu
450 455 460

Leu Gly Gln Leu Gln Asn Ile Asn Ser Leu Ile Leu Asn Asn Asn Lys
465 470 475 480

Ile His Gly Lys Ile Pro Asp Gln Leu Thr Asn Cys Phe Ser Leu Ala
485 490 495

Asn Leu Asn Ile Ser Phe Asn Asn Leu Ser Gly Ile Ile Pro Pro Met
500 505 510

Lys Asn Phe Thr Arg Phe Ser Pro Ala Ser Phe Phe Gly Asn Pro Phe
515 520 525

Leu Cys Gly Asn Trp Val Gly Ser Ile Cys Gly Pro Ser Leu Pro Lys
530 535 540

Ser Gln Val Phe Thr Arg Val Ala Val Ile Cys Met Val Leu Gly Phe
545 550 555 560

Ile Thr Leu Ile Cys Met Ile Phe Ile Ala Val Tyr Lys Ser Lys Gln
565 570 575

Gln Lys Pro Val Leu Lys Gly Ser Ser Lys Gln Pro Glu Gly Ser Thr
580 585 590

Lys Leu Val Ile Leu His Met Asp Met Ala Ile His Thr Phe Asp Asp
595 600 605

Ile Met Arg Val Thr Glu Asn Leu Asp Glu Lys Tyr Ile Ile Gly Tyr
610 615 620

Gly Ala Ser Ser Thr Val Tyr Lys Cys Thr Ser Lys Thr Ser Arg Pro
625 630 635 640

Ile Ala Ile Lys Arg Ile Tyr Asn Gln Tyr Pro Ser Asn Phe Arg Glu
645 650 655

Phe Glu Thr Glu Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Ile
660 665 670

Val Ser Leu His Gly Tyr Ala Leu Ser Pro Phe Gly Asn Leu Leu Phe
675 680 685

Tyr Asp Tyr Met Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Pro
690 695 700

Gly Lys Lys Val Lys Leu Asp Trp Glu Thr Arg Leu Lys Ile Ala Val
705 710 715 720

Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Thr Pro Arg
725 730 735

Ile Ile His Arg Asp Ile Lys Ser Ser Asn Ile Leu Leu Asp Gly Asn
740 745 750

Phe Glu Ala Arg Leu Ser Asp Phe Gly Ile Ala Lys Ser Ile Pro Ala
755 760 765

Thr Lys Thr Tyr Ala Ser Thr Tyr Val Leu Gly Thr Ile Gly Tyr Ile
770 775 780

Asp Pro Glu Tyr Ala Arg Thr Ser Arg Leu Asn Glu Lys Ser Asp Ile
785 790 795 800

Tyr Ser Phe Gly Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Ala
805 810 815

Val Asp Asn Glu Ala Asn Leu His Gln Met Ile Leu Ser Lys Ala Asp
820 825 830

Asp Asn Thr Val Met Glu Ala Val Asp Ala Glu Val Ser Val Thr Cys
835 840 845

Met Asp Ser Gly His Ile Lys Lys Thr Phe Gln Leu Ala Leu Leu Cys
850 855 860

Thr Lys Arg Asn Pro Leu Glu Arg Pro Thr Met Gln Glu Val Ser Arg
865 870 875 880

Val Leu Leu Ser Leu Val Pro Ser Pro Pro Pro Lys Lys Leu Pro Ser
885 890 895

Pro Ala Lys Val Gln Glu Gly Glu Glu Arg Arg Glu Ser His Ser Ser
900 905 910

Asp Thr Thr Thr Pro Gln Trp Phe Val Gln Phe Arg Glu Asp Ile Ser
915 920 925

Lys Ser Ser Leu
930

<210> 9

<211> 2901

<212> DNA

<213> Aradopsis thaliana ERECTA homolog

<220>

<221> CDS

<222> (1)..(2898)

<223>

<400> 9

atg aag gag aag atg cag cga atg gtt tta tct tta gca atg gtg ggt
48
Met Lys Glu Lys Met Gln Arg Met Val Leu Ser Leu Ala Met Val Gly
1 5 10 15

ttt atg gtt ttt ggt gtt gct tcg gct atg aac aac gaa ggg aaa gct
96
Phe Met Val Phe Gly Val Ala Ser Ala Met Asn Asn Glu Gly Lys Ala
20 25 30

ctg atg gcg ata aaa ggc tct ttc agc aac tta gtg aat atg ctt ttg
144

```
Leu Met Ala Ile Lys Gly Ser Phe Ser Asn Leu Val Asn Met Leu Leu
          35                  40                  45

gat tgg gac gat gtt cac aac agt gac ttg tgt tct tgg cga ggt gtt
192
Asp Trp Asp Asp Val His Asn Ser Asp Leu Cys Ser Trp Arg Gly Val
     50                  55                  60

ttc tgc gac aac gtt agc tac tcc gtt gtc tct ctg aat ttg tcc agt
240
Phe Cys Asp Asn Val Ser Tyr Ser Val Val Ser Leu Asn Leu Ser Ser
65                  70                  75                  80

ctg aat ctt gga ggg gag ata tct cca gct att gga gac cta cgg aat
288
Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Arg Asn
                85                  90                  95

ttg caa tca ata gac ttg caa ggt aat aaa cta gca ggt caa att cca
336
Leu Gln Ser Ile Asp Leu Gln Gly Asn Lys Leu Ala Gly Gln Ile Pro
           100                 105                 110

gat gag att gga aac tgt gct tct ctt gtt tat ctg gat ttg tcc gag
384
Asp Glu Ile Gly Asn Cys Ala Ser Leu Val Tyr Leu Asp Leu Ser Glu
          115                 120                 125

aat ctg tta tat gga gac ata cct ttc tca atc tct aaa ctc aag cag
432
Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln
     130                 135                 140

ctt gaa act ctg aat ctg aag aac aat cag ctc aca ggt cct gta cca
480
Leu Glu Thr Leu Asn Leu Lys Asn Asn Gln Leu Thr Gly Pro Val Pro
145                 150                 155                 160

gca acc tta acc cag att cca aac ctt aag aga ctt gat ctt gct ggc
528
Ala Thr Leu Thr Gln Ile Pro Asn Leu Lys Arg Leu Asp Leu Ala Gly
                165                 170                 175

aat cat cta acg ggt gag ata tcg aga ttg ctt tac tgg aat gaa gtt
576
Asn His Leu Thr Gly Glu Ile Ser Arg Leu Leu Tyr Trp Asn Glu Val
           180                 185                 190

ttg cag tat ctt gga tta cga ggg aat atg ttg act gga acg tta tct
624
Leu Gln Tyr Leu Gly Leu Arg Gly Asn Met Leu Thr Gly Thr Leu Ser
          195                 200                 205

tct gat atg tgt cag cta acc ggt ttg tgg tac ttt gat gtg aga gga
672
Ser Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly
     210                 215                 220
```

```
aat aat cta act gga acc atc ccg gag agc atc gga aat tgc aca agc
720
Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly Asn Cys Thr Ser
225                 230                 235                 240

ttt caa atc ctg gac ata tct tat aat cag ata aca gga gag att cct
768
Phe Gln Ile Leu Asp Ile Ser Tyr Asn Gln Ile Thr Gly Glu Ile Pro
                245                 250                 255

tac aat atc ggc ttc ctc caa gtt gct act ctg tca ctt caa gga aac
816
Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn
            260                 265                 270

aga ttg acg ggt aga att cca gaa gtt att ggt cta atg cag gct ctt
864
Arg Leu Thr Gly Arg Ile Pro Glu Val Ile Gly Leu Met Gln Ala Leu
        275                 280                 285

gct gtt ttg gat ttg agt gac aat gag ctt gtt ggt cct atc cca ccg
912
Ala Val Leu Asp Leu Ser Asp Asn Glu Leu Val Gly Pro Ile Pro Pro
    290                 295                 300

ata ctt ggc aat ctc tca ttt acc gga aag ttg tat ctc cat ggc aat
960
Ile Leu Gly Asn Leu Ser Phe Thr Gly Lys Leu Tyr Leu His Gly Asn
305                 310                 315                 320

atg ctc act ggt cca atc ccc tct gag ctt ggg aat atg tca cgt ctc
1008
Met Leu Thr Gly Pro Ile Pro Ser Glu Leu Gly Asn Met Ser Arg Leu
                325                 330                 335

agc tat ttg cag cta aac gac aat aaa cta gtg gga act att cca cct
1056
Ser Tyr Leu Gln Leu Asn Asp Asn Lys Leu Val Gly Thr Ile Pro Pro
            340                 345                 350

gag ctt gga aag ctg gag caa ttg ttt gaa ctg aat ctt gcc aac aac
1104
Glu Leu Gly Lys Leu Glu Gln Leu Phe Glu Leu Asn Leu Ala Asn Asn
        355                 360                 365

cgt tta gta ggg ccc ata cca tcc aac att agt tca tgt gca gcc ttg
1152
Arg Leu Val Gly Pro Ile Pro Ser Asn Ile Ser Ser Cys Ala Ala Leu
    370                 375                 380

aat caa ttc aat gtt cat ggg aac ctc ttg agt gga tct att cca ctg
1200
Asn Gln Phe Asn Val His Gly Asn Leu Leu Ser Gly Ser Ile Pro Leu
385                 390                 395                 400
```

```
gcg ttt cgc aat ctc ggg agc ttg act tat ctg aat ctt tcg tcg aac
1248
Ala Phe Arg Asn Leu Gly Ser Leu Thr Tyr Leu Asn Leu Ser Ser Asn
                405                 410                 415

aat ttc aag gga aaa ata cca gtt gag ctt gga cat ata atc aat ctt
1296
Asn Phe Lys Gly Lys Ile Pro Val Glu Leu Gly His Ile Ile Asn Leu
            420                 425                 430

gac aaa cta gat ctg tct ggc aat aac ttc tca ggg tct ata cca tta
1344
Asp Lys Leu Asp Leu Ser Gly Asn Asn Phe Ser Gly Ser Ile Pro Leu
        435                 440                 445

acg ctt ggc gat ctt gaa cac ctt ctc ata tta aat ctt agc aga aac
1392
Thr Leu Gly Asp Leu Glu His Leu Leu Ile Leu Asn Leu Ser Arg Asn
    450                 455                 460

cat ctt agt gga caa tta cct gca gag ttt ggg aac ctt cga agc att
1440
His Leu Ser Gly Gln Leu Pro Ala Glu Phe Gly Asn Leu Arg Ser Ile
465                 470                 475                 480

cag atg att gat gta tca ttc aat ctg ctc tcc gga gtt att cca act
1488
Gln Met Ile Asp Val Ser Phe Asn Leu Leu Ser Gly Val Ile Pro Thr
                485                 490                 495

gaa ctt ggc caa ttg cag aat tta aac tct tta ata ttg aac aac aac
1536
Glu Leu Gly Gln Leu Gln Asn Leu Asn Ser Leu Ile Leu Asn Asn Asn
            500                 505                 510

aag ctt cat ggg aaa att cca gat cag ctt acg aac tgc ttc act ctt
1584
Lys Leu His Gly Lys Ile Pro Asp Gln Leu Thr Asn Cys Phe Thr Leu
        515                 520                 525

gtc aat ctg aat gtc tcc ttc aac aat ctc tcc ggg ata gtc cca cca
1632
Val Asn Leu Asn Val Ser Phe Asn Asn Leu Ser Gly Ile Val Pro Pro
    530                 535                 540

atg aaa aac ttc tca cgt ttt gct cca gcc agc ttt gtt gga aat cca
1680
Met Lys Asn Phe Ser Arg Phe Ala Pro Ala Ser Phe Val Gly Asn Pro
545                 550                 555                 560

tat ctt tgt gga aac tgg gtt gga tct att tgt ggt cct tta ccg aaa
1728
Tyr Leu Cys Gly Asn Trp Val Gly Ser Ile Cys Gly Pro Leu Pro Lys
                565                 570                 575

tct cga gta ttc tcc aga ggt gct ttg atc tgc att gtt ctt ggc gtc
1776
```

```
Ser Arg Val Phe Ser Arg Gly Ala Leu Ile Cys Ile Val Leu Gly Val
            580                 585                 590

atc act ctc cta tgt atg att ttc ctt gca gtt tac aaa tca atg cag
1824
Ile Thr Leu Leu Cys Met Ile Phe Leu Ala Val Tyr Lys Ser Met Gln
        595                 600                 605

cag aag aag att cta caa ggc tcc tca aaa caa gct gaa ggg tta acc
1872
Gln Lys Lys Ile Leu Gln Gly Ser Ser Lys Gln Ala Glu Gly Leu Thr
    610                 615                 620

aag cta gtg att ctc cac atg gac atg gca att cat aca ttt gat gat
1920
Lys Leu Val Ile Leu His Met Asp Met Ala Ile His Thr Phe Asp Asp
625                 630                 635                 640

atc atg aga gtg act gag aat ctt aac gaa aag ttt ata att gga tat
1968
Ile Met Arg Val Thr Glu Asn Leu Asn Glu Lys Phe Ile Ile Gly Tyr
                645                 650                 655

ggt gct tct agc acg gta tac aaa tgt gca tta aaa agt tcc cga cct
2016
Gly Ala Ser Ser Thr Val Tyr Lys Cys Ala Leu Lys Ser Ser Arg Pro
            660                 665                 670

att gcc att aag cga ctc tac aat cag tat ccg cat aac ttg cgg gaa
2064
Ile Ala Ile Lys Arg Leu Tyr Asn Gln Tyr Pro His Asn Leu Arg Glu
        675                 680                 685

ttt gag aca gaa ctt gag acc att ggg agc att agg cac aga aac ata
2112
Phe Glu Thr Glu Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Ile
    690                 695                 700

gtc agc ttg cat gga tat gcc ttg tct cct act ggc aac ctt ctt ttc
2160
Val Ser Leu His Gly Tyr Ala Leu Ser Pro Thr Gly Asn Leu Leu Phe
705                 710                 715                 720

tat gac tac atg gaa aat gga tca ctt tgg gac ctt ctt cat ggg tca
2208
Tyr Asp Tyr Met Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Ser
                725                 730                 735

ttg aag aaa gtg aag ctt gat tgg gag aca agg ttg aag ata gcg gtt
2256
Leu Lys Lys Val Lys Leu Asp Trp Glu Thr Arg Leu Lys Ile Ala Val
            740                 745                 750

gga gct gca caa gga cta gcc tat ctt cac cac gat tgt act cct cga
2304
Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Thr Pro Arg
        755                 760                 765
```

```
atc att cac cgt gac atc aag tca tcg aac ata ctt ctt gat gag aat
2352
Ile Ile His Arg Asp Ile Lys Ser Ser Asn Ile Leu Leu Asp Glu Asn
    770                 775                 780

ttc gaa gca cat tta tct gat ttc ggg att gct aag agc ata cca gct
2400
Phe Glu Ala His Leu Ser Asp Phe Gly Ile Ala Lys Ser Ile Pro Ala
785                 790                 795                 800

agc aaa acc cat gcc tcg act tat gtt ttg gga aca att ggt tat ata
2448
Ser Lys Thr His Ala Ser Thr Tyr Val Leu Gly Thr Ile Gly Tyr Ile
                805                 810                 815

gac cca gag tat gct cgt act tca cga atc aat gag aaa tcc gat ata
2496
Asp Pro Glu Tyr Ala Arg Thr Ser Arg Ile Asn Glu Lys Ser Asp Ile
            820                 825                 830

tac agc ttc ggt att gtt ctt ctt gag ctt ctc act ggg aag aaa gca
2544
Tyr Ser Phe Gly Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Ala
        835                 840                 845

gtg gat aac gaa gct aac ttg cat caa ctg ata ttg tca aag gct gat
2592
Val Asp Asn Glu Ala Asn Leu His Gln Leu Ile Leu Ser Lys Ala Asp
    850                 855                 860

gat aat act gtg atg gaa gca gtt gat cca gag gtt act gtg act tgt
2640
Asp Asn Thr Val Met Glu Ala Val Asp Pro Glu Val Thr Val Thr Cys
865                 870                 875                 880

atg gac ttg gga cat atc agg aag aca ttt cag ctg gct ctc tta tgc
2688
Met Asp Leu Gly His Ile Arg Lys Thr Phe Gln Leu Ala Leu Leu Cys
                885                 890                 895

aca aag cga aac cct tta gag aga ccc aca atg ctt gaa gtc tct agg
2736
Thr Lys Arg Asn Pro Leu Glu Arg Pro Thr Met Leu Glu Val Ser Arg
            900                 905                 910

gtt ctg ctc tct ctt gtc cca tct ctg caa gta gca aag aag cta cct
2784
Val Leu Leu Ser Leu Val Pro Ser Leu Gln Val Ala Lys Lys Leu Pro
        915                 920                 925

tct ctt gat cac tca acc aaa aag ctg cag caa gag aat gaa gtt agg
2832
Ser Leu Asp His Ser Thr Lys Lys Leu Gln Gln Glu Asn Glu Val Arg
    930                 935                 940
```

aat cct gat gca gaa gca tct caa tgg ttt gtt cag ttc cgt gaa gtc
2880
Asn Pro Asp Ala Glu Ala Ser Gln Trp Phe Val Gln Phe Arg Glu Val
945 950 955 960

atc tcc aaa agt agc ata taa
2901
Ile Ser Lys Ser Ser Ile
965

<210> 10

<211> 966

<212> PRT

<213> Aradopsis thaliana ERECTA homolog

<400> 10

Met Lys Glu Lys Met Gln Arg Met Val Leu Ser Leu Ala Met Val Gly
1 5 10 15

Phe Met Val Phe Gly Val Ala Ser Ala Met Asn Asn Glu Gly Lys Ala
20 25 30

Leu Met Ala Ile Lys Gly Ser Phe Ser Asn Leu Val Asn Met Leu Leu
35 40 45

Asp Trp Asp Asp Val His Asn Ser Asp Leu Cys Ser Trp Arg Gly Val
50 55 60

Phe Cys Asp Asn Val Ser Tyr Ser Val Val Ser Leu Asn Leu Ser Ser
65 70 75 80

Leu Asn Leu Gly Gly Glu Ile Ser Pro Ala Ile Gly Asp Leu Arg Asn
85 90 95

Leu Gln Ser Ile Asp Leu Gln Gly Asn Lys Leu Ala Gly Gln Ile Pro
100 105 110

Asp Glu Ile Gly Asn Cys Ala Ser Leu Val Tyr Leu Asp Leu Ser Glu
115 120 125

Asn Leu Leu Tyr Gly Asp Ile Pro Phe Ser Ile Ser Lys Leu Lys Gln
130 135 140

```
Leu Glu Thr Leu Asn Leu Lys Asn Asn Gln Leu Thr Gly Pro Val Pro
145                 150                 155                 160

Ala Thr Leu Thr Gln Ile Pro Asn Leu Lys Arg Leu Asp Leu Ala Gly
                165                 170                 175

Asn His Leu Thr Gly Glu Ile Ser Arg Leu Leu Tyr Trp Asn Glu Val
            180                 185                 190

Leu Gln Tyr Leu Gly Leu Arg Gly Asn Met Leu Thr Gly Thr Leu Ser
        195                 200                 205

Ser Asp Met Cys Gln Leu Thr Gly Leu Trp Tyr Phe Asp Val Arg Gly
    210                 215                 220

Asn Asn Leu Thr Gly Thr Ile Pro Glu Ser Ile Gly Asn Cys Thr Ser
225                 230                 235                 240

Phe Gln Ile Leu Asp Ile Ser Tyr Asn Gln Ile Thr Gly Glu Ile Pro
                245                 250                 255

Tyr Asn Ile Gly Phe Leu Gln Val Ala Thr Leu Ser Leu Gln Gly Asn
            260                 265                 270

Arg Leu Thr Gly Arg Ile Pro Glu Val Ile Gly Leu Met Gln Ala Leu
        275                 280                 285

Ala Val Leu Asp Leu Ser Asp Asn Glu Leu Val Gly Pro Ile Pro Pro
    290                 295                 300

Ile Leu Gly Asn Leu Ser Phe Thr Gly Lys Leu Tyr Leu His Gly Asn
305                 310                 315                 320

Met Leu Thr Gly Pro Ile Pro Ser Glu Leu Gly Asn Met Ser Arg Leu
                325                 330                 335

Ser Tyr Leu Gln Leu Asn Asp Asn Lys Leu Val Gly Thr Ile Pro Pro
            340                 345                 350

Glu Leu Gly Lys Leu Glu Gln Leu Phe Glu Leu Asn Leu Ala Asn Asn
        355                 360                 365
```

```
Arg Leu Val Gly Pro Ile Pro Ser Asn Ile Ser Ser Cys Ala Ala Leu
370                 375                 380

Asn Gln Phe Asn Val His Gly Asn Leu Leu Ser Gly Ser Ile Pro Leu
385                 390                 395                 400

Ala Phe Arg Asn Leu Gly Ser Leu Thr Tyr Leu Asn Leu Ser Ser Asn
                405                 410                 415

Asn Phe Lys Gly Lys Ile Pro Val Glu Leu Gly His Ile Ile Asn Leu
            420                 425                 430

Asp Lys Leu Asp Leu Ser Gly Asn Asn Phe Ser Gly Ser Ile Pro Leu
        435                 440                 445

Thr Leu Gly Asp Leu Glu His Leu Leu Ile Leu Asn Leu Ser Arg Asn
    450                 455                 460

His Leu Ser Gly Gln Leu Pro Ala Glu Phe Gly Asn Leu Arg Ser Ile
465                 470                 475                 480

Gln Met Ile Asp Val Ser Phe Asn Leu Leu Ser Gly Val Ile Pro Thr
                485                 490                 495

Glu Leu Gly Gln Leu Gln Asn Leu Asn Ser Leu Ile Leu Asn Asn Asn
            500                 505                 510

Lys Leu His Gly Lys Ile Pro Asp Gln Leu Thr Asn Cys Phe Thr Leu
        515                 520                 525

Val Asn Leu Asn Val Ser Phe Asn Asn Leu Ser Gly Ile Val Pro Pro
    530                 535                 540

Met Lys Asn Phe Ser Arg Phe Ala Pro Ala Ser Phe Val Gly Asn Pro
545                 550                 555                 560

Tyr Leu Cys Gly Asn Trp Val Gly Ser Ile Cys Gly Pro Leu Pro Lys
                565                 570                 575

Ser Arg Val Phe Ser Arg Gly Ala Leu Ile Cys Ile Val Leu Gly Val
            580                 585                 590
```

Ile Thr Leu Leu Cys Met Ile Phe Leu Ala Val Tyr Lys Ser Met Gln
595 600 605

Gln Lys Lys Ile Leu Gln Gly Ser Ser Lys Gln Ala Glu Gly Leu Thr
610 615 620

Lys Leu Val Ile Leu His Met Asp Met Ala Ile His Thr Phe Asp Asp
625 630 635 640

Ile Met Arg Val Thr Glu Asn Leu Asn Glu Lys Phe Ile Ile Gly Tyr
645 650 655

Gly Ala Ser Ser Thr Val Tyr Lys Cys Ala Leu Lys Ser Ser Arg Pro
660 665 670

Ile Ala Ile Lys Arg Leu Tyr Asn Gln Tyr Pro His Asn Leu Arg Glu
675 680 685

Phe Glu Thr Glu Leu Glu Thr Ile Gly Ser Ile Arg His Arg Asn Ile
690 695 700

Val Ser Leu His Gly Tyr Ala Leu Ser Pro Thr Gly Asn Leu Leu Phe
705 710 715 720

Tyr Asp Tyr Met Glu Asn Gly Ser Leu Trp Asp Leu Leu His Gly Ser
725 730 735

Leu Lys Lys Val Lys Leu Asp Trp Glu Thr Arg Leu Lys Ile Ala Val
740 745 750

Gly Ala Ala Gln Gly Leu Ala Tyr Leu His His Asp Cys Thr Pro Arg
755 760 765

Ile Ile His Arg Asp Ile Lys Ser Ser Asn Ile Leu Leu Asp Glu Asn
770 775 780

Phe Glu Ala His Leu Ser Asp Phe Gly Ile Ala Lys Ser Ile Pro Ala
785 790 795 800

Ser Lys Thr His Ala Ser Thr Tyr Val Leu Gly Thr Ile Gly Tyr Ile
805 810 815

Asp Pro Glu Tyr Ala Arg Thr Ser Arg Ile Asn Glu Lys Ser Asp Ile

820 825 830

Tyr Ser Phe Gly Ile Val Leu Leu Glu Leu Leu Thr Gly Lys Lys Ala
835 840 845

Val Asp Asn Glu Ala Asn Leu His Gln Leu Ile Leu Ser Lys Ala Asp
850 855 860

Asp Asn Thr Val Met Glu Ala Val Asp Pro Glu Val Thr Val Thr Cys
865 870 875 880

Met Asp Leu Gly His Ile Arg Lys Thr Phe Gln Leu Ala Leu Leu Cys
885 890 895

Thr Lys Arg Asn Pro Leu Glu Arg Pro Thr Met Leu Glu Val Ser Arg
900 905 910

Val Leu Leu Ser Leu Val Pro Ser Leu Gln Val Ala Lys Lys Leu Pro
915 920 925

Ser Leu Asp His Ser Thr Lys Lys Leu Gln Gln Glu Asn Glu Val Arg
930 935 940

Asn Pro Asp Ala Glu Ala Ser Gln Trp Phe Val Gln Phe Arg Glu Val
945 950 955 960

Ile Ser Lys Ser Ser Ile
965

Dated this SECOND day of JULY, 2002

The Australian National University
Patent Attorneys for the Applicant:

F B RICE & CO

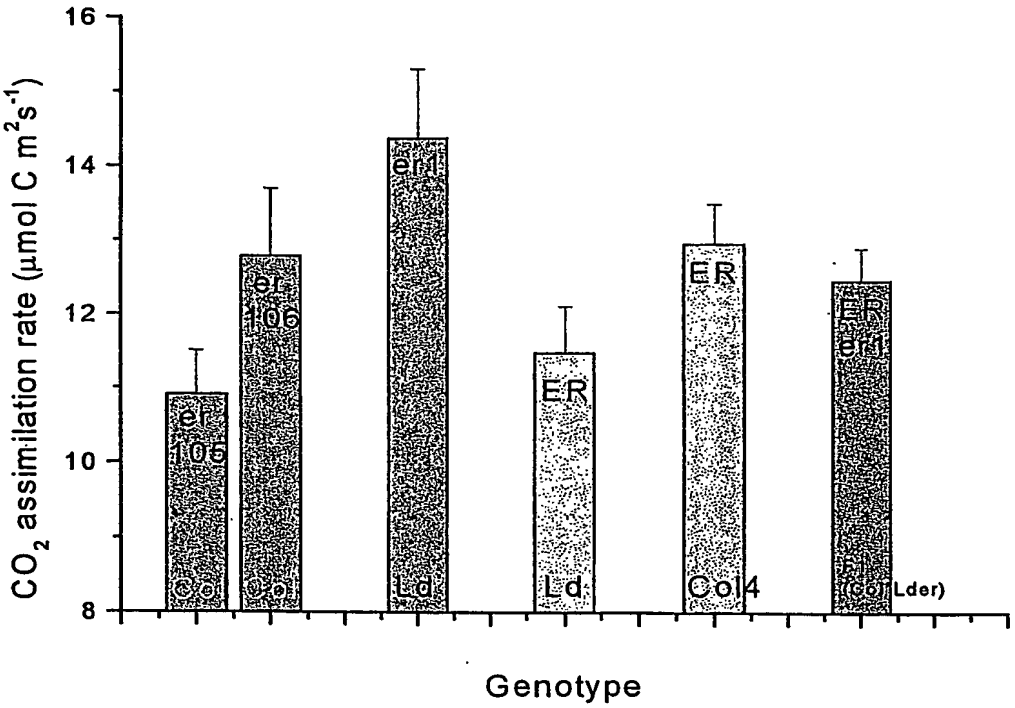
16
14
12
10
8
CO_2 assimilation rate (µmol C m^2s^{-1})
er 105
Col
er 105
Col
er1
Ld
ER
Ld
ER
Col4
ER er1
F1 (Col Lder)
Genotype

Figure 1a

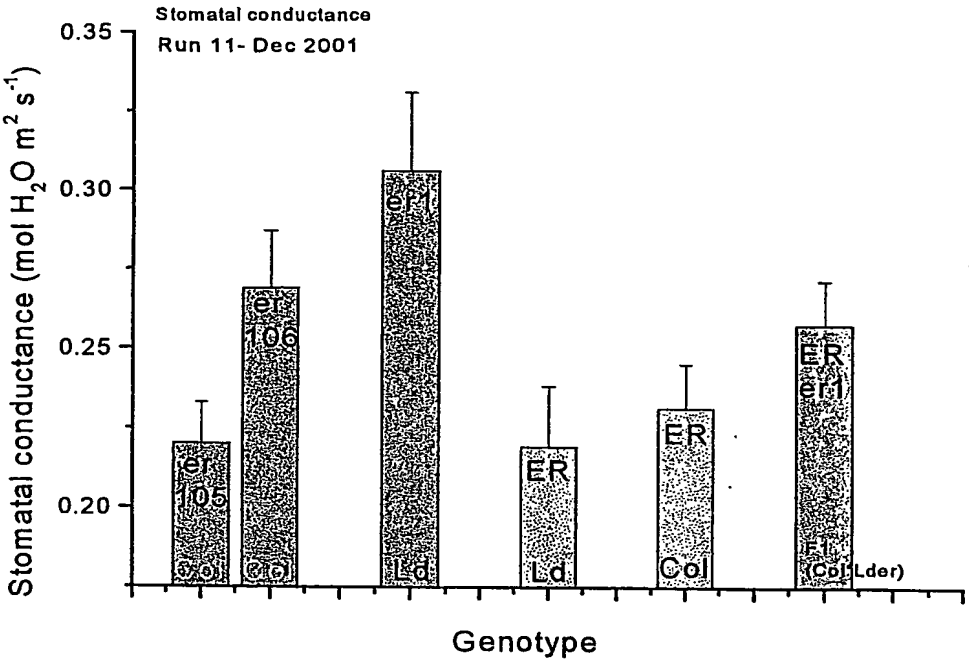
Stomatal conductance
Run 11- Dec 2001
Stomatal conductance (mol H_2O m^2 s^{-1})
0.35
0.30
0.25
0.20
er 105
er 106
eF1
ER
ER
ER er-1
F1 (Col Lder)
Col
Ld
Ld
Genotype

Figure 1b

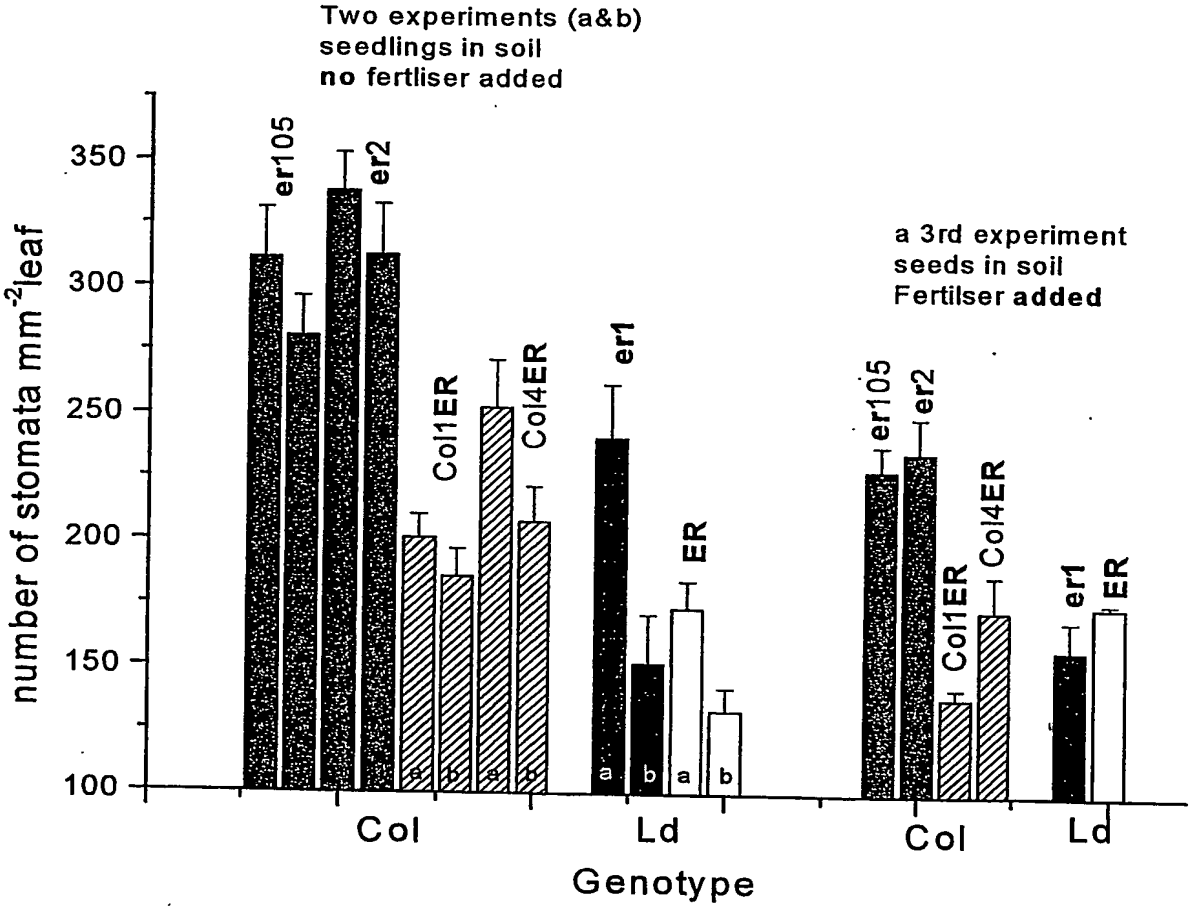
Two experiments (a&b)
seedlings in soil
no fertliser added
a 3rd experiment
seeds in soil
Fertilser added
number of stomata mm-2leaf
350
300
250
200
150
100
er105
er2
Col1ER
Col4ER
er1
ER
a
b
Col
Ld
Genotype

Figure 2a

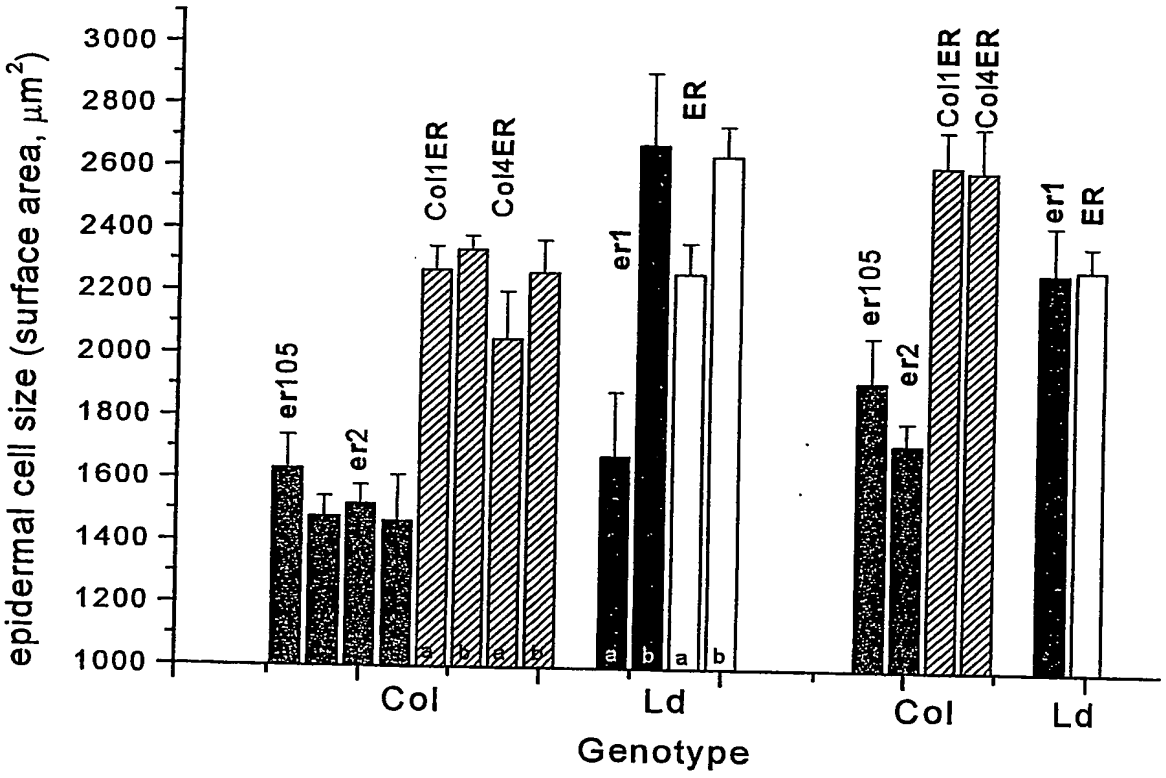
epidermal cell size (surface area, µm²)
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
er105
er2
Col1ER
Col4ER
er1
ER
er105
er2
Col1ER
Col4ER
er1
ER
a
b
a
b
a
b
a
b
Col
Ld
Col
Ld
Genotype

Figure 2b

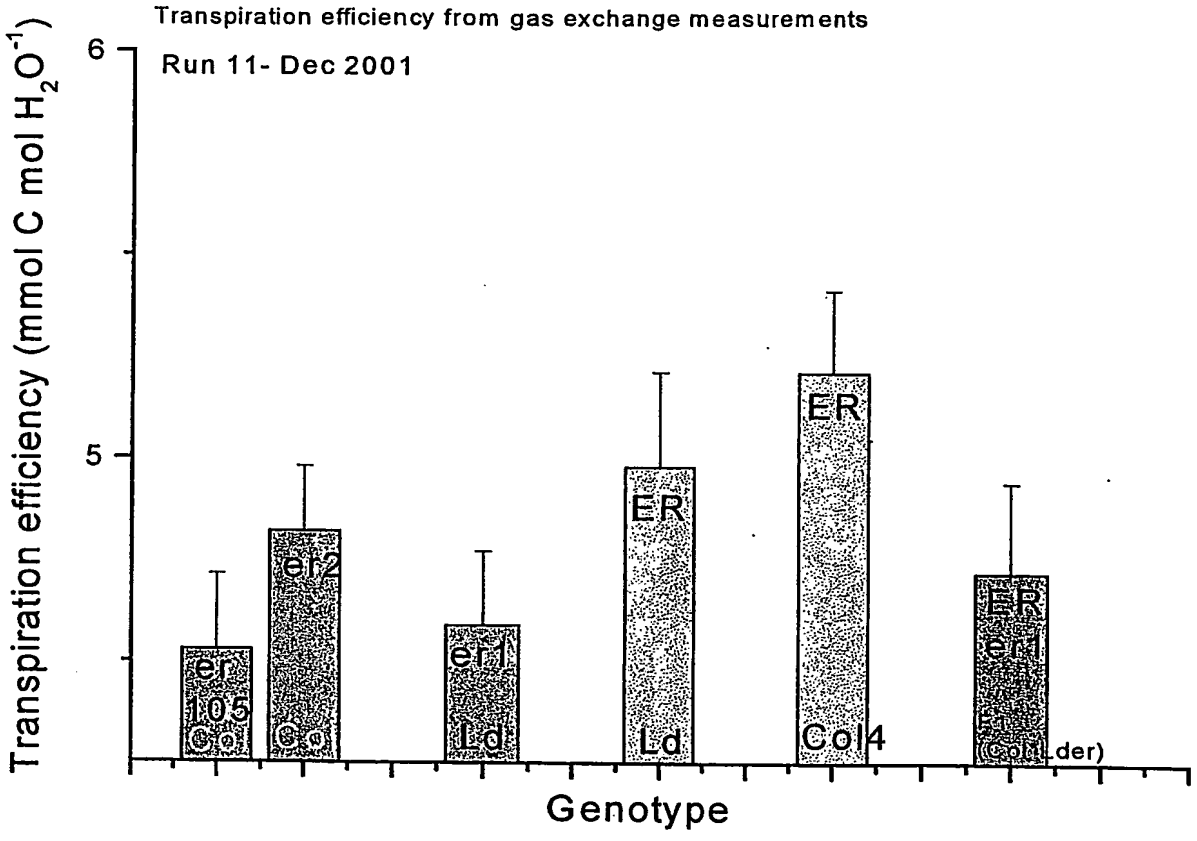
Transpiration efficiency from gas exchange measurements
Run 11- Dec 2001
Transpiration efficiency (mmol C mol H_2O^{-1})
6
5
er
105
Co
er2
Co
er1
Ld
ER
Ld
ER
Col4
ER
er1
Col der)
Genotype

Figure 1c

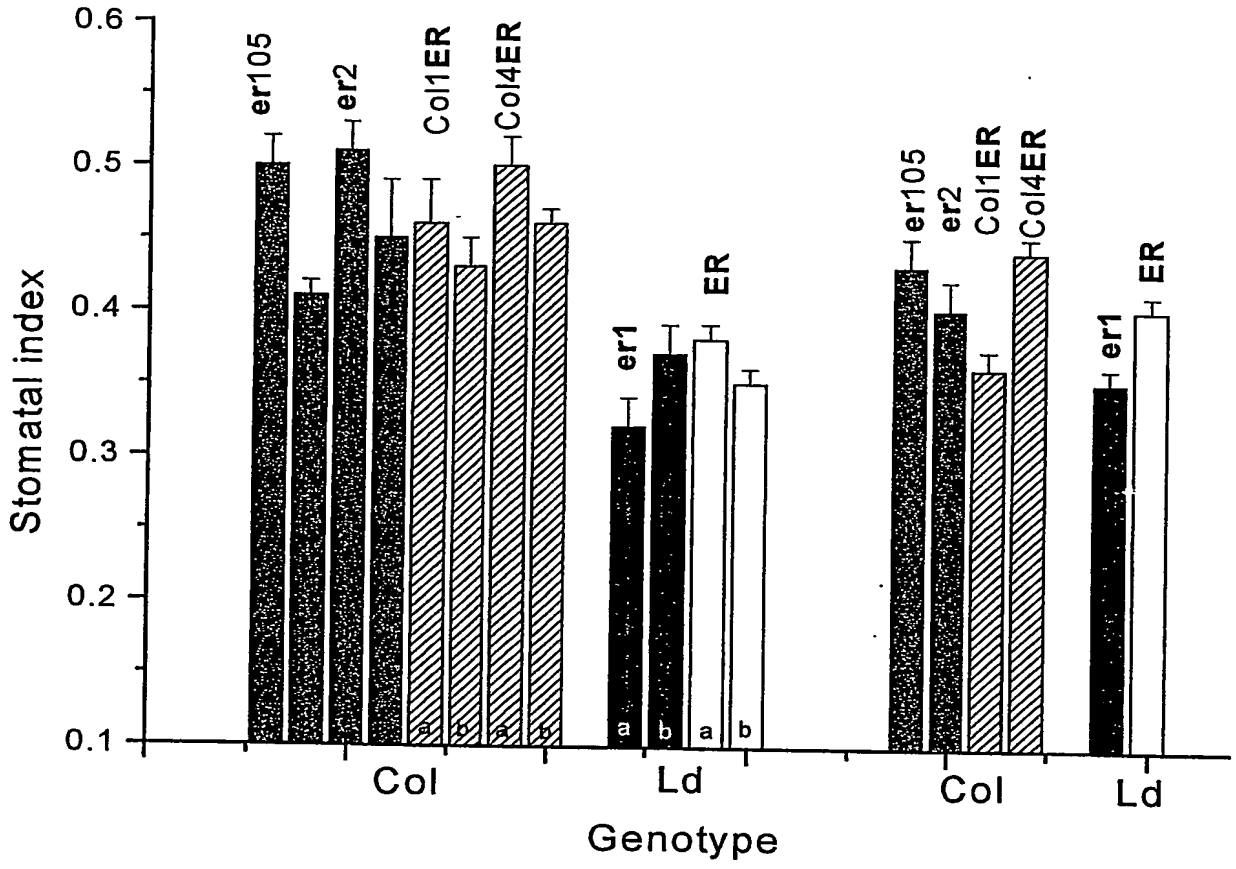
0.6
0.5
0.4
0.3
0.2
0.1
Stomatal index
er105
er2
Col1ER
Col4ER
er1
ER
a
b
a
b
a
b
a
b
Col
Ld
Genotype
er105
er2
Col1ER
Col4ER
er1
ER
Col
Ld

Figure 2c